Evidence and Practice: How Surgeons Know Their Operations Work

David Jones, MD, PhD
Harvard Medical School

Introduction:
This module is designed to be a guided discussion, not a lecture presentation. It does not intend to convey historical facts but to demonstrate how history can be used to foster thoughtful analysis of surgery. The materials provide an outline for the session, questions for discussion, and background material that the moderator can provide via slides to help prompt the discussion. While the focus of the module is on a specific procedure (coronary artery bypass grafting), there are many opportunities for the discussion to take up other procedures, depending on student and moderator interest and expertise. The material could be covered briefly (20 minutes), but more time would allow fuller discussions of the issues it raises.

Goals:
The module uses a historical case-study -- the debates about randomized trials of coronary artery bypass grafting between 1968 and 1978 -- to raise important questions for surgeons. How do surgeons know that their treatments work? How have standards for judging efficacy changed over time? Why have there been fierce debates about research methodology? What does this reveal about surgery and how surgeons think?

Background:
Surgery, more so than other forms of medical intervention, allows multiple ways of assessing therapeutic outcomes. Surgeons can judge whether a procedure:

(1) alleviates the patient's symptoms
(2) produces the desired rearrangement of tissues and organs
(3) restores a physiological function
(4) improves a measurable finding, whether laboratory or radiologic
(5) prevents a future outcome
(6) improves long-term survival

The efficacy of an operation often seems self-evident (e.g., the pain is gone, the mass was removed, the obstruction was relieved, the coronary blood flow has been increased, etc.).

But many key outcomes, especially improved long-term survival or decreased need for subsequent procedures (e.g., repeat revascularization), cannot be demonstrated so easily. When can the efficacy of a procedure be assumed? When does it require extensive and rigorous clinical testing? Surgeons should ask these questions of every new -- and existing -- procedure.

Surgical procedures have other features that raise the stakes of these questions. While the invasiveness of surgical interventions often makes their efficacy seem obvious (e.g., the gangrenous limb was amputated), this invasiveness, because of its potential morbidity and mortality, demands that surgeons submit their operations to strict scrutiny. Since surgical procedures are sometimes tried as last-ditch efforts in patients who have no other options (e.g., metastatic cancer, refractory coronary disease), the hope that a procedure might help can be more important for patients than robust evidence that the procedure is likely to help. And yet false hope can be devastating for both patients and their families.

Over the course of the twentieth century, the methods of clinical research have become increasingly sophisticated, from single case reports or case series in the early twentieth century, to multi-center prospective randomized trials, sham-controls, or registry studies today. Surgeons have employed all of these methods. However, the introduction of each new method provoked fierce controversy as surgeons raised legitimate concerns about the limitations of the new methods. The nature of the debates that have embroiled randomization, blinding, or sham controls reveals much about surgeons and surgery. Students should emerge from the discussion with a more nuanced way of thinking about what kinds of operations require what degree of scrutiny and methodological rigor.

Plan:
The module is designed for the moderator to review the materials in advance. When the session begins, the moderator can show each slide, read the text in italics, and start a discussion with the suggested questions. However, it is not necessary to adhere closely to the scripts.

Slide 1: Introduction

The history of surgery offers many lessons for surgeons today. History can show not just how surgical practice has changed over time, but also how surgical thinking has changed, and continues to change. One key question is what kinds of clinical research are needed to demonstrate the efficacy of surgical procedures. This module uses a historical case study to examine why answers to this question have been interesting and controversial.

Slide 2: How Do We Know Surgery Works?

The history of surgery has many examples of operations, once popular, that have been discredited. Bloodletting was used for thousands of years before falling out of favor in the 19th and early 20th centuries. Radical mastectomy was used widely from the 1880s until the 1970s. Prefrontal lobotomy earned its inventor the 1949 Nobel Prize. Debates continue today about many procedures, including the management of DCIS or the choice of surgery vs. antibiotics for appendicitis. This history shows that, for some procedures, surgeons' judgments about efficacy and appropriateness have changed over time.

Ask the students a series of initial questions. How do surgeons know if treatments work? What methods exist for demonstrating the efficacy of surgical procedures? Do different operations require different degrees of scrutiny and methodological rigor?

If needed, the moderator can prompt the students with specific examples. Do some procedures have self-evident efficacy, e.g., suture closure of simple skin lacerations? There are important questions about even such simple procedures Has suture closure actually been tested in a randomized trial? What studies have been done of sutures vs. steri-strips vs. skin glue? Have studies examined outcomes of laceration closure by medical students vs. interns vs. experienced plastic surgeons? Have any of the students actually read a published study about this procedure, a procedure that they presumably have done many times?

Other procedures raise much more difficult questions, e.g., gender reassignment surgery. What kinds of outcomes should surgeons consider for a case like this? Surgeons have looked at proximate outcomes (e.g., cosmetic appearance, sexual function), intermediate outcomes (e.g., patient satisfaction, ability to engage in fulfilling sexual encounters), and long-term outcomes, including overall quality of life, mental health, and even suicide rates.

Slide 3: The Emergence of Coronary Artery Bypass Grafting

Coronary artery disease has been the leading cause of death in the United States since 1910 (with the exception of the 1918 flu pandemic). It is now the leading cause of death worldwide. From its rise to prominence in the 1910s to 1930s, until the 1960s, it was mostly managed medically, often with unsatisfying results (in the absence of ß-blockers, ACE inhibitors, statins, etc.). Surgeons developed many surgical procedures that sought to reduce demand on the heart (e.g., sympathectomy or thyroidectomy) or improve blood flow to the heart (e.g., creating adhesions between the heart and other vascularized tissues, implanting the internal thoracic artery into the myocardium). Although surgeons were enthusiastic about some of these procedures (e.g., a 1966 NEJM *study about myocardial implants), few of these procedures were done (e.g., fewer than 1000 per year). CABG, which had been proposed in 1910, and studied in animal models since the 1950s, was first used clinically in 1960. It received significant attention -- and enthusiasm -- with the work of René Favaloro and Donald Effler at the Cleveland Clinic in 1968.*

Ask the students: When CABG was first done, how could (should?) surgeons have determined whether it worked? What are the possible goals of surgery in a disease like CAD? There are many possibilities (e.g., restoration of blood flow, relief of angina, prolonged survival, others); what study designs are needed to producing convincing evidence for these different goals?

Slide 4: The Initial Efficacy Claims

The early reports of CABG described the efficacy of the procedure in different ways. Robert Goetz wrote in an addendum to an animal study that his first patient was "doing well"; he planned to write a detailed case report, but never did. René Favaloro began a series of saphenous vein interposition grafts; in the 15th patient in this series, the entire proximal RCA was occluded, so he performed his first aorta-coronary bypass graft. He reported that all the patients survived the procedure and that postoperative angiography showed good blood flow through the grafts. A follow up study of their first 100 patients reported one intra-operative death, four post-operative deaths, and three late deaths. In the 50 patients studied with post-operative angiography, the grafts were "patent without narrowing" in 39, and narrowed less than 50% in another; these 40 patients were "free of angina." Favaloro's chief, Donald Effler, wrote than any cardiologist unconvinced by such evidence was "allowing emotion to prevail over scientific evaluation." As surgeons published ever-larger case series to demonstrate the efficacy of CABG, the operation became one of the most frequently performed major surgical procedures in the United States, with over 100,000 procedures per year by 1977.

Ask the students: Is this adequate evidence of efficacy? Why or why not? Can a single case report or a 100-patient case series produce convincing evidence of improved

coronary flow?

If the goals of CABG include preventing heart attacks and improving survival, what kinds of evidence and study design are required to demonstrate efficacy? Can you assume that if blood flow has been restored, then survival will be improved?

Slide 5: Cardiologists' Critiques

The rapid spread of CABG fueled a dramatic reaction from cardiologists. While some enthusiastically referred their patients to surgeons for the procedures, others demanded that surgeons produce more convincing evidence of efficacy. CABG skeptics had many concerns. Outcomes varied significantly between surgical programs (e.g., operative mortality varied from <4% to >11%). It was not yet clear how long the grafts would remain patent. Publications often had either no control group, or relied on comparisons to historical controls. This made it difficult to compare the results of surgery to the results of optimized medical therapy, especially since medical therapy was also in flux (e.g., the introduction of ß-blockers). Prominent cardiologists and outcomes researchers demanded that surgeons perform a randomized trial of CABG.

Ask the students to brainstorm about what objections surgeons might have had to randomized trials. They should be able to develop a long list, e.g., the self-evident efficacy of the procedure, variations in operator skill, inability to blind the patients, time lag (i.e., by the time trials results become available, the procedure will have been refined and improved). It does not matter that they provide an exhaustive list, or the right list.

Slide 6: The VA Cooperative Study and Surgeons' Responses

The first major randomized trial of CABG -- and one of the first major randomized trials of any surgical procedure -- was conducted by a consortium of Veterans Administration hospitals and published in the New England Journal of Medicine *in 1977. The study found no significant difference in survival rates at 36 months, except in the patients with the most severe forms of CAD. Surgeons were quick to highlight the limitations of the trial; they published scathing critiques in many journals. The VA team and other CABG skeptics responded with their own critiques of the critiques.*

Ask the students to offer their own assessment of the study and its critiques:

• Should CABG utilization have decreased in response to the study? The study may have had a brief impact, but CABG rates grew steadily from the late 1970s into the mid 1990s. Enthusiasm for CABG more or less survived the trial intact. Elite surgeons believed that CABG would have performed if they had done the operations themselves. Even when

surgeons accepted the mortality data, they reminded that CABG produced dramatic relief of angina and that that alone justified the procedure.

• Would it have been better if the country's elite cardiac surgeons had performed the trial, instead of the VA surgeons? Why didn't they?

• Why was the debate so ferocious? The management of millions of patients was at stake. With surgeons and hospitals receiving high fees for the operation (surgeon's fees could exceed $10,000, and elite surgeons performed many CABG procedures each week), the financial stakes were enormous.

Ask the students to design an ideal trial of CABG. Would such a trial be feasible? Many trials have been done since the 1970s (e.g., CABG vs. medications; CABG vs. PCI; etc.). For the most part, they sustain the basic findings of the VA Study: CABG provides a mortality benefit only in the sickest patients.

Ask the students to generalize beyond CABG. For any surgical procedure, what are the key challenges to randomized trials? In what ways are randomized trials of surgical procedures more challenging than those of medications?

Slide 7: The Inevitable Limits of Randomized Trials in Surgery

Even if you could design an optimal trial, there will always be limitations. One problem is operator skill: different surgeons can have different skill levels, and this affects surgical outcomes. Should a trial be done by the elite surgeons, to demonstrate the potential of a procedure, or by typical surgeons, to reveal what surgical outcomes will be like in the community? There are also significant learning curves. Should randomized trials be done early in the history of a procedure, before entrenched enthusiasm sets in, or only once surgeons have optimized and mastered the new procedure? A second problem is time lag. For a procedure like CABG, where the ideal outcome is long-term survival, it can take 5 to 10 years between study design and publication of the results. If surgeons have continued to innovate in the meantime, the study might simply end up reporting the results of an outdated procedure. A third problem is blinding. While medication trials can be blinded easily, surgical trials are difficult to blind, both practically and ethically. The control procedure (e.g., a sham operation) needs to be indistinguishable from the actual operation, and this requires anesthesia, incisions, etc., and their attendant risks.

Ask the students whether it is worthwhile to do randomized trials of surgery, given these limitations -- is a flawed "gold standard" still useful? Is it ever ethical to use sham surgical controls to test the efficacy of a surgical procedure? If so, when, and why?

Slide 8: Sham Controlled Surgical Trials
Blinded studies of surgical procedures have been done, especially in orthopedics. Surgeons, for instance, have tested arthroscopic knee surgery for osteoarthritis, vertebroplasty for osteoporotic spinal fractures, and partial meniscectomy for degenerative meniscal tears. The researchers went to great lengths to insure that the patients in the procedure and control groups had indistinguishable operating room experiences. In each of these three cases, the "active" surgical procedure performed no better than the sham "placebo" control. These negative findings, about popular surgical procedures, demonstrate the potential value of sham controlled trials, despite the ethical controversy about them.

Ask the students whether they think sham control trials should be used more widely in surgery. Why or why not?

There is some selection bias in these three studies: surgeons have only been willing to do sham controlled trials when there has been significant skepticism about the value of the procedure. Few recommend sham trials of procedures for which there is significant enthusiasm. While this makes sense, is it wise?

Slide 9: What Does It Take to Provide Convincing Evidence of Efficacy?
Over the past fifty years, standards of surgical research have changed dramatically. Randomized trials, despite continuing concerns about their inherent limitations, have become common in surgery. However, other modes of knowledge production exist side by side. Case series continue to be published, and can be useful in some cases. Registry studies are become increasingly more valuable, as the depth and breadth of registry data improve, and as researchers develop new methods of analyzing registry data. Surgeons must make deliberate decisions about why kinds of studies to do in order to produce convincing data about surgical efficacy.

Ask the students to brainstorm to develop guidelines about what kind of data might be needed for different kinds of procedures. How is it different for operations that treat medical emergencies vs. chronic symptoms? Are there markers of therapeutic success that are evident immediately (allowing short duration trials), or is long-term data needed? How well is the pathophysiology and the natural history of the disease understood? In the case of CABG, for instance, there has been confusion about whether one short term outcome -- relief of angina -- correlates with longer term actuarial outcomes -- prevention of myocardial infarction and prolongation of survival. Can students develop a "sliding

scale" in which studies of different kinds of operations would require different degrees of methodological rigor? There is no right answer here: this remains an area of ongoing debate.

Slide 10: Conclusions and Further Reading

Throughout your careers, you will encounter many new procedures. Surgeons will make efficacy claims based on a wide variety of evidence. You will need to become savvy consumers of these claims and develop sophisticated standards for judging whether or not convincing evidence of efficacy has been produced. Even procedures with seemingly self-evident efficacy, such as CABG, clearly deserve strict scrutiny.

Robert H. Goetz, Michael Rohman, Jordan D. Haller, Ronald Dee, and Stephan S. Rosenak, "Internal Mammary-Coronary Artery Anastomosis: A Nonsuture Method Employing Tantalum Rings," *Journal of Thoracic and Cardiovascular Surgery* 41 (1961): 378-386.

Richard Gorlin and Warren J. Taylor, "Selective Revascularization of the Myocardium by Internal-Mammary-Artery Implant," *New England Journal of Medicine* 275 (1966): 283-290.

René G. Favaloro, "Saphenous Vein Autograft Replacement of Severe Segmental Coronary Artery Occlusion," *Annals of Thoracic Surgery* 5 (1968): 334-339.

René G. Favaloro, Donald B. Effler, Laurence K. Groves, William C. Sheldon, and Mohammed Riahi, "Direct Myocardial Revascularization with Saphenous Vein Autograft: Clinical Experience in 100 Cases," *Diseases of the Chest* 56 (1969): 279-283.

Donald B. Effler, "The Role of Surgery in the Treatment of Coronary Artery Disease," *Annals of Thoracic Surgery* 8 (1969): 376-379.

Eugene Braunwald, "Direct Coronary Revascularization ... A Plea Not To Let the Genie Escape from the Bottle," *Hospital Practice* 6 (1971): 9-10.

Thomas C. Chalmers, "Randomization and Coronary Artery Surgery," *Annals of Thoracic Surgery* 14 (1972): 323-327.

David H. Spodick, "Numerators Without Denominators: There is No FDA for the Surgeon," *JAMA* 232 (1975): 35-36.

Jack W. Love, "In Comment: Drugs and Operations: Some Important Differences," *JAMA* 232 (1975): 37-38.

Timothy Takaro, Herbert N. Hultgren, Martin J. Lipton, Katherine M. Detre, and Participants in the Study Group, "The VA Cooperative Randomized Study of Surgery for Coronary Arterial Occlusive Disease: II. Subgroup with Significant Left Main Lesions," *Cardiovascular Surgery* 54 suppl 3 (1976): 107-117.

Marvin L. Murphy, Herbert N. Hultgren, Katherine Detre, James Thomsen, Timothy

Takaro, and Participants of the Veterans Administration Cooperative Study, "Treatment of Chronic Stable Angina: A Preliminary Report of Survival Data of the Randomized Veterans Administration Cooperative Study," *New England Journal of Medicine* 297 (1977): 621-627.

Michael E. DeBakey and Gerald M. Lawrie, "Aortocoronary-Artery Bypass," *JAMA* 239 (1978): 837-839.

David S. Jones, "Visions of a Cure: Visualization, Clinical Trials, and Controversies in Cardiac Therapeutics, 1968-1998," *Isis* 91 (2000): 504-541.

J. Bruce Moseley, K.O. O'Malley, N.J. Petersen, T.J. Menke, B. Brody, B., D.H. Kaykendall, J.C. Hollingsworth, C.M. Ashton, and N.P. Wray, "A Controlled Trial of Arthroscopic Surgery for Osteoarthritis of the Knee," *New England Journal of Medicine* 347 (2002): 81-88.

Rita F. Redberg, "Sham Controls in Medical Device Trials," *New England Journal of Medicine* 371 (2014): 892-893.

David S. Jones, "The Puzzle of Positive Results — Myocardial Revascularization," *New England Journal of Medicine* 372 (2015): 501-503.