Improving Perioperative Outcomes: My Journey into Risk, Patient Preferences, Guidelines, and Performance Measures

Ninth Honorary FAER Research Lecture

Lee A. Fleisher, M.D.*

As a clinical anesthesiologist and health services researcher, my goal is to improve patient outcomes both on an individual and a population level through direct patient care and research. In recent years, I have had the privilege of being involved in the development of numerous guidelines.1–7 Guidelines are a set of recommendations for patient management that identifies a specific or range of management strategies based on the evidences. Guidelines are usually developed and promulgated by a specialty medical society or government agency (i.e., National Institutes of Health or Agency for Healthcare Research and Quality). It is important to recognize that guidelines are the culmination of both research and expert opinion and that application of guidelines can improve care, but if done incorrectly, it can lead to worse outcomes. It is within this context that I would like to review my perspective on the research that underlies both guidelines and their implementation into clinical practice.

The goal of improving perioperative outcomes comes under the general rubric of outcomes research. The agency for healthcare research and quality defines outcomes research as:

- The study of understanding of the end results of particular healthcare practices and interventions.
- End results include effects that people experience and care about, such as change in the ability to function, in addition to more traditional measures such as death and major morbidity.

When focusing on improving outcomes in anesthesiology, it is important to define those outcomes of interest including death; major organ dysfunctions, such as myocardial infarction, pneumonia, and renal failure; and minor morbidity, including pain, nausea, and vomiting. There is also increasing interest in patient-oriented outcomes such as quality of recovery, quality of life, and patient satisfaction. The time horizon of interest to the anesthesiologist has traditionally been 24 to 48 h, although several investigators have suggested that perioperative actions by the anesthesiologist may impact outcomes at 1 yr or longer.8

In thinking about the influence of anesthesia on perioperative risk, it is important to recognize that there is a risk that is directly attributable to anesthesia (e.g., loss of airway), whereas there are outcomes related to the stress of undergoing surgery, which can be mitigated by the care provided by an anesthesiologist. With respect to the former, anesthesiologists have made major advances in making anesthesia safer. It is estimated that the risk of death directly attributable to anesthesia may be in the order of 1 death in 185,000 anesthetics, although one investigator has questioned this conclusion.9,10 However, perioperative organ dysfunction is common after surgery, and it is associated with an increased cost of care.11 As perioperative physicians, our goal should be to mitigate perioperative and long-term organ dysfunction by application of the best evidence.

One approach to achieve this goal is to begin by assessing or determining those patients at greatest risk for major and minor morbidity and mortality. Once the patients at risk are identified, it is important to study their actions or interventions that might influence risk and reduce complications. If an intervention is found to be either efficacious or effective (the difference is discussed later) then it is important to disseminate this information. If the level of evidence is sufficiently strong or there is a preponderance of expert opinion on a topic, then a guideline or practice parameter can be

* Robert D. Dripps Professor of Anesthesiology and Critical Care and Professor of Medicine, Department of Anesthesiology and Critical Care, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania.

Received from Department of Anesthesiology and Critical Care, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania. Submitted for publication November 1, 2009. Accepted for publication November 4, 2009. Supported in part by the National Institutes of Health (Bethesda, Maryland) R01 DK073671-01A1. Presented in part as the Foundation for Anesthesia Education and Research Honorary Research Lecture 2009 at the American Society of Anesthesiologists Annual Meeting in New Orleans, Louisiana, October 19, 2009. The figures in this article were prepared by Dimitri Karetnikov, 7 Tennyson Drive, Plainsboro, New Jersey 08536.

Address correspondence to Dr. Fleisher: Department of Anesthesiology and Critical Care, University of Pennsylvania School of Medicine, 6 Dulles, Philadelphia, Pennsylvania 19104. lee.fleisher@uphs.upenn.edu. This article may be accessed for personal use at no charge through the Journal Web site, www.anesthesiology.org.
developed to better disseminate the information to the individual practitioner. However, applying the evidence to the individual patient must be done within the framework of understanding the benefits and risks associated with the intervention and how the individual patient views the potential trade-offs from differing interventions and outcomes. Finally, it is important to recognize that even if evidence is properly disseminated, the adoption of new and important evidence may be slow. One of the paradigms that the current government has adopted to increase the speed of diffusion and adoption is the use of performance measures. However, the value of measuring and paying for “correct” performance leads to improved outcomes is a matter of debate. This article attempts to review the quest of anesthesiologists for improving perioperative and periprocedure outcomes within this proposed framework.

Understanding the Risks and Benefits of an Intervention

Identifying Risk

An underlying assumption in medical care is that treatment should be individualized for a given patient if optimal outcomes are desired. In the perioperative period, anesthesiologists have multiple options with regard to the choice of anesthesia, monitors, and drugs used to achieve the desired outcome. A large portion of our patients will do well if they receive a standard anesthetic (e.g., propofol, a muscle relaxant, an inhalational agent, and narcotic). There is a basic assumption that obtaining a preoperative history, physical examination, and appropriate laboratory tests will influence management choices, which will then lead to better outcomes. The concept of establishing a baseline database to assess risk of disease and risk of developing complications can be framed within a decision paradigm (fig. 1). By using our understanding of clinical findings, previous experience, and the application of the literature, we can assess the probability of disease within certain confidence intervals (CI). Within this paradigm, the clinician may decide that the probability of disease reaches some threshold for action. For patients with cardiovascular disease who are undergoing noncardiac surgery, these actions can include modification of medical management including initiation or continuation of β-blockers and statins, treatment for unstable coronary symptoms, and coronary revascularization among other interventions.4

Therefore, risk assessment is a critical first step from both the clinical and research perspectives. One of the first and best known studies of assessing cardiovascular risk is the Cardiac Risk Index published by Goldman et al.12 During the ensuing three decades, others have further refined the Index, and most recently, Lee et al.13 published the revised Cardiac Risk Index. They identified six risk factors with an increasing probability of developing cardiac events with an increasing number of factors present. Important to any such development is the validation of these indices, and many such studies have been published recently.

Another major area of interest of my group is admission after outpatient surgery. Given the low rates of admission, administrative date (e.g., state discharge summaries and Medicare claims data) has been used to determine predictors of admission or death.14,15 An outpatient surgery admission index was created that establishes the risk factors, and based on the number of risk factors, the probability of admission after outpatient surgery was defined.15 Based on this probability, I believe that the patient and the physician can determine the optimal location of care (free standing vs. hospital-based outpatient surgery) to yield the best outcome.

Importantly, the clinical evaluation and integration of knowledge does not always lead to a clear decision for action. In some cases, the probability of disease is so high that the optimal action is well defined because the probability is far above any threshold for action. Alternatively, if patients have only minor risk factors for coronary disease then they clearly do not require any further evaluation because no action would be taken. Similarly, patients with known and severe coronary artery disease who are already in their optimal medical condition also rarely need further evaluation because they are above the threshold for action but the action is likely already taken. A key question is the optimal action when the probability and extent of disease is less precise, and the threshold for any action lies within the CIs of the risk assessment. Testing can potentially have value in such situations because a negative test would decrease the probability and obviate any need for action, whereas a positive test would increase the probability above threshold and therefore lead to the action.

Within this framework, cardiovascular testing has undergone an evolution, during the past two decades, from initially being used routinely in all patients to being used selectively based on an assessment of risk. In addition, I and others have helped to refine interpretation of the test. For example, a larger effect or area of regional wall motion abnormality denotes...
greater risk in dipyridamole thallium imaging and dobutamine stress echocardiography when compared with a simple positive or negative scan. This approach formed the basis of guidelines for perioperative cardiovascular evaluation.

**Understanding the Evidence**

Randomized clinical trials form the strongest basis for determining the evidence supporting an action, and during the past several decades, there has been a marked increase in their number and quality in the perioperative period. Randomized clinical trials are designed to prove efficacy in determining whether an intervention works under ideal conditions. They have defined patient inclusion/exclusion criteria and usually have strict protocols of care. Although the internal validity of these trials is high, the external validity of the trials may be low. However, effectiveness refers to how an intervention works under real world conditions. In such situations, the intervention may behave identical or different from the randomized clinical trial.

In the patient with cardiovascular disease, examples of randomized clinical trials include the study of β-blockers, statins, coronary revascularization, and thermal management. The recent series of studies of perioperative β-blockade in noncardiac surgery help illustrate both the importance of how the baseline risks of the patient and how study protocol can influence effectiveness. For example, β-blockade started several weeks in advance of surgery has been shown to be efficacious in vascular surgery patients at high risk. However, subsequent studies questioned the finding in patients at lower risk and in real world usage (i.e., effectiveness).

To study efficacy in lower risk populations, a larger sample size may be required. In DECREASE IV, β-blockers were studied in a randomized trial of 1,066 patients at intermediate risk. Patients randomized to bisoprolol had a lower incidence of perioperative cardiac death and nonfatal myocardial infarction than those randomized to bisoprolol-control (2.1% vs. 6.0% events; hazard ratios: 0.34; 95% CI 0.17–0.67). Importantly, bisoprolol was started at least 7 days in advance of surgery. To study the effectiveness in low-risk patients, administrative datasets (e.g., Medicare claims data) may prove useful. For example, Lindemann et al. retrospectively reviewed the records of 782,969 patients and determined who received β-blocker treatment during the first 2 days of hospital stay. The relationship between perioperative β-blocker treatment and the risk of death varied directly with cardiac risk. Among the 580,665 patients with a revised Cardiac Risk Index score of 0 or 1, treatment was associated with no benefit and possible harm, whereas among the patients with a revised Cardiac Risk Index score of 2, 3, and 4 or more, the adjusted odds ratios for death in the hospital were 0.88 (95% CI 0.80–0.98), 0.71 (95% CI 0.63–0.80), and 0.58 (95% CI 0.50–0.67), respectively (fig. 2).

The Perioperative Ischemic Evaluation Study (POISE) trial demonstrated the importance of study protocol with respect to the effectiveness of a drug. A total of 8,351 patients were randomized to controlled-release oral metoprolol succinate or placebo. The primary endpoint of cardiac death, nonfatal myocardial infarction, or cardiac arrest was reduced in the metoprolol group compared with placebo (5.8% vs. 6.9%; hazard ratio 0.84; 95% CI 0.70–0.99; P = 0.04), driven by a reduction of nonfatal myocardial infarctions, however, at the costs of an increased incidence of total mortality and stroke. One of the major conjectures with regard to the difference between the results in the Perioperative Ischemic Evaluation Study and other trials is the differences in the protocol. Metoprolol succinate, a long-acting β-blocker, was used at much higher doses and was started on the morning of the surgery in contrast to lower doses titrated to effect and started weeks before surgery.

**Adoption or Diffusion of New Information**

As new knowledge is acquired, it is important to disseminate the information to the individual practitioner. Presentations at national meetings and publication of studies in journals are the two primary modes of dissemination. Adoption and diffusion of innovation (such as procedures, drugs, and devices) have been well studied. Rogers proposed an adoption or diffusion model based on agricultural studies. Initially, there is innovation followed by the early adopters. As seen in figure 3, this is followed by the early majority, the late majority, and finally the laggards. The innovators and early adopters can be used as change agents and influence others to adopt best practices. However, the overall speed of adoption may be slow. It is well known that many interventions that are found to be effective take years to be adopted. For example, thrombolytic therapy was not advocated in textbooks for nearly 20 yr after it was shown to be effective in a meta-analytic analysis (fig. 4).

**How Do We Increase the Speed of Adoption: Guidelines**

Although there are large number of articles published annually, the ability to keep abreast of all of the new knowledge...
and to integrate the information into a coherent practice is difficult. Traditionally, review articles by experts offered a means of integrating knowledge, but these articles have many limitations. For example, the authors may not approach the review of the literature in a systematic way, and there may be selection bias in the studies included in the analysis. Therefore, these articles frequently include the bias of the author. During the past two decades, there has been a movement toward evidence-based medicine and the use of systematic approaches to the review of the literature and the study of medical interventions. Some of the techniques used in evidence-based medicine include systematic reviews, meta-analysis, and cost-effectiveness analysis.27

One means of codifying these evidence-based reviews is through the development of standards and guidelines. It is important to recognize that a standard implies that a therapy or practice should be used for patients with a particular condition. Standards are only approved if an assessment of the probabilities and uses of the group indicate that the decision to choose the treatment or a strategy would be virtually unanimous and should be applied rigidly, although many standards do not have a body of literature to support their use. One such example is pulse oximetry monitoring for anesthesia.28 However, guidelines are a set of recommendations for patient management that identifies a specific or range of management strategies based on the evidence. Guidelines are usually developed and promulgated by a specialty medical society or government agency (i.e., National Institutes of Health or Agency for Healthcare Research and Quality). Formal recommendations are published, which include the strength of recommendation. For example, the American College of Cardiology/American Heart Association classifications of evidence that are used in their Guidelines to summarize indications for a particular therapy or treatment are as follows:

- Class I: Conditions for which there is evidence or general agreement that the procedure or therapy is useful and effective.
- Class II: Conditions for which there is conflicting evidence or a divergence of opinion about the usefulness or efficacy of performing the procedure or therapy.
- Class IIa: Weight of evidence or opinion is in favor of usefulness or efficacy.
- Class IIb: Usefulness or efficacy is less well established by evidence or opinion.
- Class III: Conditions for which there is evidence or general agreement that the procedure or therapy is not useful or effective, and in some cases, it may be harmful.

The levels of evidence are as follows:

- Level of evidence A: recommendation based on evidence from multiple randomized trials or meta-analyses.
• Level of evidence B: recommendation based on evidence from a single randomized trial or nonrandomized studies.
• Level of evidence C: recommendation based on expert opinion, case studies, or standards of care.

The problem is that the quality of the evidence, even in the well-studied area of cardiovascular disease, is poor. Tricoci et al. evaluated 16 current American College of Cardiology/American Heart Association Guidelines reporting levels of evidence and found that only 314 recommendations of 2,711 total are classified as level of evidence A, whereas 1,246 are level of evidence C. Recommendations with the level of evidence A are mostly concentrated in class I, but only 245 of 1,305 class I recommendations has level of evidence A.

Implementation of Recommendations

Implementation to the Individual Patient

If an intervention is found to be effective or advocated in a Guideline, then should it always be implemented in every patient? Clearly, the answer is no because the patient of interest may not be similar to the original population studied or the processes of care, and the method of implementation is different at the local institution than those of the original studies. As described earlier, patients have different levels of risk, and in addition, patients have a complex characteristic of diseases for which the best practice may actually be in conflict. Boyd et al. demonstrated that most clinical practice guidelines do not modify or discuss the applicability of their recommendations for older patients with multiple comorbidities. They also demonstrated that adverse interactions between drugs and diseases could result. Therefore, applying best practices (which may be defined in guidelines) requires an understanding of the applicability of the results and the risks and benefits in the individual patient. This is a further application of efficacy versus effectiveness and illustrates the importance of incorporating this information into guidelines and having defined triggers for reevaluating the evidence.

Patient preferences are an important component of the application of evidence to the individual patient. Specifically, different patients value and assign different weights to different outcomes, and therefore, the optimal decision for any given patient is sensitive to these values. Patient preferences can be assessed using willingness-to-pay, which has been applied to issues such as nausea, vomiting, and choice of inpatient or outpatient care. Another approach to assess preferences is the use of standard gamble techniques, whereby the subject is offered a choice of a known outcome and also offered a varying probabilities of a good outcome of morbidity or mortality. An example of this technique is used on the television show "Deal or No Deal" in which the contestant is offered a defined dollar amount from the bank versus a chance that their case contains a higher or lower amount.

These techniques have frequently been applied to determine how subjects rate such outcomes as stroke, angina, and so forth, and reflect the risk taking or risk aversion of the subject. My research group has applied these techniques to perioperative questions and demonstrated that for mild to moderate pain and for nausea and vomiting, most patients would prefer going home after laparoscopic cholecystectomy, although the percentage of patients willing to take the risk of going home is much smaller for severe pain.

A critical area in which patient preferences impact on decision-making is in the area of decisions regarding preoperative cardiovascular testing in patients undergoing intermediate risk surgery with one to two risk factors. As outlined in the American College of Cardiology/American Heart Association Guidelines, proceeding to the operating room with heart rate control was assigned a class IIa recommendation, while preoperative testing was assigned a class IIb recommendation. In trying to determine the correct decision for the individual patient, a patient’s preference for surgery versus alternative management can drive the decision. For example, if the patient is determined to be at high risk, then they may choose to undergo a less-invasive procedure or no surgery at all.

Performance Measures

President Obama’s drive to reform healthcare delivery and to reduce its costs has focused interest on value-based purchasing. The Centers for Medicare and Medicaid Services describes value-based purchasing as “a strategy that can help to transform the current payment system by rewarding providers for delivering high quality, efficient clinical care.” The optimal strategy would be to pay hospitals and providers based on their outcome; however, the importance of baseline risk in assessment on outcome is well known. Despite numerous attempts at defining optimal risk-adjustment methodologies, none have been adopted for payment purpose. Another means of achieving the goal of rewarding outcome is through the development of best-practice performance measures (practices known to lead to improved outcome) and linking service payments to achieve the performance goals. The underlying assumption is that physicians do not adopt best practices outlined in the literature or in Guidelines, as documented by the low initial compliance rates on practices such as appropriate antibiotic timing, continuation of β-blockers, and venous thromboembolism prophylaxis shown in the Surgical Care Improvement Project.

There are two general types of performance measures: hospital based and physician based. Examples of hospital-based measures are those advocated by Surgical Care Improvement Project. There is currently no actual link between performance and payment in the Medicare program but rather between reporting and payment. Rates of compliance with the measure can be found on the Hospital compare website.†

Physician level performance measures are also used by the Medicare program and by many private insurance firms. These measures are often developed by the specialty societies

(Committee for Performance and Outcome Measures of the American Society of Anesthesiologists) and endorsed by the American Medical Association or other national bodies and eventually the National Quality Forum. Perioperative examples include administration of antibiotics in a timely manner and maintenance of normothermia.

It is critical to ensure that such performance measures are truly linked to the evidence and that they change with changes in the evidences. For example, there were numerous groups, including Leapfrog, which initially set the standard for β-blocker therapy as acute administration in all patients at risk. The Surgical Care Improvement Project technical expert panel was concerned about the unpublished results of the POISE trial, and therefore chose to only include continuation of β-blockers in those patients already taking these agents as the measure until the trial was published.

Another concern is the potential unintended consequence of adopting practices uniformly. For example, many of the measures of delivering antibiotics within a specified window in the emergency department for suspected pneumonia may lead to overtreatment of patients who have heart failure. In addition, there is a great deal of concern that the goal should not be 100% compliance because compliance with preoperative antibiotic timing for surgical procedures may lead to multiple preoperative doses. Among hospitals participating in a voluntary quality-improvement initiative, the pay-for-performance program was not associated with a significant incremental improvement in quality of care or outcomes for acute myocardial infarction.34 Importantly, the investigators did not find evidence that pay for performance had an adverse association with improvement in processes of acute myocardial infarction care that were not subject to financial incentives.

There are other concerns about paying for performance. Some investigators have questioned the value of paying for performance as opposed to voluntary quality improvement programs. Lindenauer et al.35 evaluated adherence to 10 individuals and four composite measures of quality during a period of 2 yr at 613 hospitals that voluntarily reported information about the quality of care through a national public reporting initiative, including 207 facilities that simultaneously participated in a pay-for-performance demonstration project funded by the Centers for Medicare and Medicaid Services. They reported that compared with the control group, pay-for-performance hospitals showed greater improvement in all composite measures of quality, including measures of care for heart failure, acute myocardial infarction, and pneumonia and a composite of 10 measures. After adjustments were made for differences in baseline performance and other hospital characteristics, pay-for-performance was associated with improvements ranging from 2.6 to 4.1% during the 2-yr period.

Most importantly, it is unclear whether pay-for-performance initiatives actually lead to improve outcome. Werner and Bradlow36 performed a cross-sectional study of hospital care between January 1 and December 31, 2004, for acute myocardial infarction, heart failure, and pneumonia at acute care hospitals in the United States included in the Hospital Compare Web site and compared them with hospital risk-adjusted mortality rates, which were measured using Medicare Part A claims data. Across all acute myocardial infarction performance measures, the absolute reduction in risk-adjusted mortality rates between hospitals performing in the 25th percentile versus those performing in the 75th percentile was 0.005 for inpatient mortality, 0.006 for 30-day mortality, and 0.012 for 1-yr mortality. Differences in mortality rates for hospitals performing in the 75th percentile on all measures within a condition versus those performing lower than the 25th percentile on all reported measures for acute myocardial infarction ranged between 0.008 \((P = 0.06)\) and 0.018 \((P = 0.008)\). They concluded that hospital performance measures predict small differences in hospital risk-adjusted mortality rates and that effort should be made to develop performance measures that are tightly linked to patient outcomes.

Finally, if organ dysfunction and complications were reduced, it is unclear whether mortality would also decrease.

---

**Fig. 5.** Rates of all complications, major complications, and death after major complications, according to hospital quintile of mortality. Reproduced with permission from *N Engl J Med* 2009; 361:1368–75.

Lee A. Fleisher

Anesthesiology, V 112 • No 4 • April 2010
Silber et al. proposed the concept of “failure to rescue” as a measure of hospital quality. These investigators suggested that measuring mortality in hospitalized patients who developed complications varies dramatically. Factors associated with improved failure to rescue include higher nursing staff ratios and the presence of board-certified anesthesiologists. Much of this research was developed using administrative data, specifically Medicare claims. Using validated 30-day mortality and morbidity attained from the American College of Surgeons National Surgical Quality Improvement Program, investigators demonstrated that the risk-adjusted rate of complications was similar between hospitals enrolled in the program but mortality varied twofold, and therefore, failure to rescue varied twofold in this population (fig. 5). Therefore, reducing the rate of complications, an important goal, may not reduce 30-day mortality.

Summary
During the past several decades, there has been a concerted effort to improve perioperative outcomes through the performance of randomized controlled trials and dissemination and synthesis of the data into recommendations of care. Anesthesiologists can have significant impact on perioperative outcome through the application of evidence to reduce organ dysfunction. To achieve those goals, it is critical that anesthesiologists apply the evidence and help design processes to be appropriate to the individual patient. Despite these questions regarding the current approach to pay-for-performance, it is clear that quality improvement programs that report outcomes do result in improvements. It is also evident that anesthesiologists can reduce the rate of failure-to-rescue, and therefore can impact the rate of death in patients who develop complications. We must strive to continue to improve the outcomes of our patients through a willingness to measure ourselves.

References
tion and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Circulation 1999; 100:1043–9
36. Werner RM, Bradlow ET: Relationship between Medicare’s hospital compare performance measures and mortality rates. JAMA 2006; 296: 2694–702