Linking Processes and Outcomes to Improve Surgical Performance

A New Approach to Morbidity and Mortality Peer Review

Michael Pine, MD, MBA,* Donald E. Fry, MD, FACS†

ABSTRACT

Peer review of surgical cases resulting in death or potentially avoidable complications is a longstanding tradition, but intensive reviews of individual cases rarely produce tangible improvements in clinical outcomes. The systematic comparison of care received by patients who experienced adverse outcomes to care received by patients who had uneventful surgical courses is a promising alternative to intensive case review. However, results of these studies may be misleading because physicians often choose interventions based on their perceptions of patients’ preoperative risks, and higher adverse outcome rates among patients with higher preoperative risk may distort comparisons of alternative interventions. The creation of a control sample by carefully matching each patient who experienced an adverse outcome to a patient who had a similar preoperative risk but did not experience an adverse outcome can overcome this problem and provide excellent insight into how to improve clinical performance most effectively and efficiently. By using currently available electronic data to compute each patient’s risk of an adverse outcome, a series of cases with adverse outcomes can be matched to an equal number of controls. Peer review committees can then direct focused data collection and analyses of potentially critical processes of care to determine which, if any, are associated with significantly poorer clinical results. A simulated scenario is analyzed in detail to illustrate the ability of this technique to correctly determine best practices when other approaches fail to do so.

Introduction

Peer review of surgical cases resulting in death or potentially avoidable complications is a longstanding tradition. This tradition is embodied in the Joint Commission on Accreditation of Healthcare Organization’s requirement that formal reviews of mortality and morbidity be conducted at all accredited hospitals.1 In general, these reviews focus on deciding whether bad outcomes in individual cases were the unavoidable consequences of patients’ underlying conditions, were possibly preventable if care had been optimal rather than ordinary, or were the direct consequences of substandard care. The medical literature (sometimes embodied in clinical practice guidelines) and individual professional experience provide standards by which the quality of surgical care is evaluated. Often a patient’s outcome influences assessment of the quality of surgical care so that more severe complications resulting in permanent injury are more likely to be attributed to substandard care than are transient, less severe complications.2-3

Although surgical judgment and technique are the focus of most mortality and morbidity reviews, increasing attention is being given to adverse outcomes as sentinel events arising from flaws in systems of delivering care.4 The latter approach emphasizes the importance of well-designed, properly managed systems in achieving success in complex endeavors that require the coordinated efforts of many individuals. The importance of surgical policies and procedures in ensuring good clinical outcomes is supported by studies of complex manufacturing processes in which more than 80% of quality problems were attributable to faulty systems rather than to substandard individual performance.5 In practice, it is even more difficult to determine objectively which, if any, policies and/or procedures were responsible for a single adverse outcome than it is to determine whether failure to adhere to a clinical guideline was an acceptable clinical decision associated with an unfortunate outcome or an error in judgment that resulted in injury or death.

Intensive review of single cases to determine the cause of adverse outcomes may identify instances of substandard practice and may, at times, be instrumental in improving clinical practice through physician education and institutional sanctions. Case reviews also may identify suboptimal processes of care and catalyze improvements in policies and procedures that affect important clinical outcomes. But for all the effort expended on surgical peer review, individual surgical
practices and organizational policies and procedures rarely are 
affected by these deliberations. Some of this failure to achieve 
constructive change is the natural product of organizational 
inertia. On the other hand, resistance to change often flows 
from uncertainty about whether proposed changes will 
address identified clinical problems successfully. The absence 
of a rigorous scientific method of assessing the relation 
between each of the myriad of factors that may contribute to 
an adverse outcome and the occurrence of the adverse out-
come in individual patients severely limits the utility of peer 
review as a vehicle for quality improvement.

The following hypothetical scenario created using simulated 
data illustrates a technique for relating processes of care to clinical 
outcomes in the context of surgical peer review. It illustrates 
how differences in the treatment of high- and low-risk 
patients can result in incorrect inferences about what consti-
tutes good surgical practice. It also demonstrates how data 
retrieval, data analysis, and clinical judgment can be combined 
most efficiently to revitalize surgical peer review and transform 
it into an effective instrument for clinical and organizational 
quality improvement.

**Simulated Illustrated Case Study: Part 1**

Dr. Gordon Lightouch, the Chief of Surgery at Cookbook 
Memorial Medical Center, is convinced that using general 
anesthesia for patients undergoing lower extremity reconstruc-
tion increases mortality substantially and that regional anes-
thesia always should be used in these cases. Doctor Philip 
Outcolde, the Chief of Anesthesia, strongly disagrees and 
believes that the choice between a general and a regional anes-
thetic should be left to the anesthetist. During peer review of a 
patient who died two days after lower extremity reconstruc-
tion with general anesthesia, Doctor Lightouch cited several 
articles in highly respected surgical journals that favored 
regional anesthesia for these operations. Doctor Outcolde 
counteracted by citing several studies published in highly respect-
ed anesthesia journals, all of which supported his contention 
that the type of anesthetic agent administered during these 
operations did not affect the occurrence of adverse outcomes. 
Doctor Outcolde also suggested that failure to adhere to the 
hospital’s protocol for antibiotic prophylaxis in this case was a 
more likely cause of the patient’s death than was the use of a 
general anesthetic. Doctor Lightouch accused Doctor Out-
colde of obfuscation, and the quality of the discussion deterio-
rated rapidly.

Gwen Provit, Director of Quality Improvement at Cook-
book Memorial, shook her head sadly. It appeared to her that 
this was the millionth time she had heard this argument 
between Doctor Lightouch and Doctor Outcolde. Worse still, 
the two protagonists appeared to be no closer to resolving their 
differences than they were the first time these issues were 
raised. Provit knew of successful programs that strongly 
favored regional anesthesia and successful programs that used 
general and regional anesthesia almost interchangeably. She 
knew of successful programs that strictly adhered to antibiotic 
protocols and successful programs that had not even adopted 
such protocols. It was clear that what worked well in one envi-
ronment might not work at all in another. The critical ques-
tion was how to manage surgical cases at Cookbook Memorial 
to produce the best possible clinical outcomes.

To answer this question, Provit analyzed 600 consecutive 
cases of lower extremity reconstruction performed at Cook-
book General. Of these, 540 had been discharged alive and 
60 had died in the hospital; 480 received a general anesthet-
ic and 120 received a regional anesthetic; 300 received pro-
phylactic antibiotics according to protocol and 300 did not. 
Provit reasoned that practices that were clinically efficacious 
should be associated with good outcomes (in this case, sur-
vival), and practices that were ill-advised should be associat-
ed with poor outcomes (in this case, death). However, 
instead of grouping patients based on their treatment, as is 
done in conventional clinical trials to determine the efficacy 
of an intervention, she grouped patients according to their 
outcome. This permitted her to evaluate the effect of multiple 
potentially influential interventions (in this case, anes-
thesia and antibiotics) on a single outcome of interest (in this case, death).

Because reliable information about specific potentially 
fluential interventions often is unavailable in standard elec-
tronic records, this information must be abstracted directly 
from individual patients’ paper medical records. Because the 
cost of this type of data retrieval can be prohibitive, it is impor-
tant to minimize the amount of data that must be collected in 
this manner. To do this, Provit limited her analysis to the 60 
patients who died in the hospital and a random sample consist-
ing of 60 of the 540 patients who were discharged alive.

Of the 60 patients who died, 55 had received a general anes-
thetic and 5 had received a regional anesthetic. In the absence 
of a control group of patients who survived, these data certain-
ly would favor the avoidance of general anesthesia in lower 
extrremity reconstructions. However, of the 60 patients who 
survived, 40 had received a general anesthetic and 20 had 
received a regional anesthetic. Thus, the introduction of a con-
trol group into the analysis weakens but does not completely 
undermine the case for avoiding general anesthesia. Extrapo-
lating back to the entire population, these data suggest that 
general anesthesia was preferred for this operation, with 415 
patients receiving a general anesthetic (ie, 55 fatalities and 40 
× 9 = 360 survivors) and 185 patients receiving a regional anes-
thetic (ie, 5 fatalities and 20 × 9 = 180 survivors). This analysis 
demonstrates the danger of focusing only on cases with adverse 
outcomes. In the absence of data about the treatment of sur-
vivors, one could conclude that anesthesia was responsible for 
all 60 deaths because all of these patients received either a gen-
eral or a regional anesthetic.

To determine the probability that the difference in the use 
of a general anesthetic between fatalities and survivors was 
due to chance alone (ie, random variation), Provit performed 
a y-square test. If the type of anesthetic administered had no 
effect on survival, 47.5 patients who died and 47.5 patients 
who lived would have received a general anesthetic, and 12.5
patients who died and 12.5 patients who lived would have received a regional anesthetic. The absolute values of the differences between the observed numbers and these expected numbers all are 7.5; the calculated χ²-square statistic is 11.37 with 1 degree of freedom which is greater than 6.63, the value associated with a 1% probability of the difference being due to chance alone. Therefore, Provit concluded that death was related to the use of a general anesthetic and that the probability that this association was due to chance alone was less than 1%.

Provit repeated this analysis using data about adherence to the Medical Center’s antibiotic protocol for this procedure. Of the 60 patients who died, 27 were treated according to the protocol and 33 were not. Of the 60 patients who survived, 26 were treated according to the protocol and 34 were not. Extrapolating back to the complete data set, these data suggest that the antibiotic protocol was followed about 45% of the time, with 261 patients treated according to the protocol (ie, 27 fatalities and 26 x 9 = 234 survivors) and 339 patients not treated according to the protocol (ie, 33 fatalities and 34 x 9 = 306 survivors).

To determine the probability that the difference in adherence to the antibiotic protocol between fatalities and survivors was due to chance alone, Provit performed a second χ²-square test. If adherence to the protocol had no effect on survival, 26.5 patients who died and 26.5 patients who lived would have been treated according to the protocol, and 33.5 patients who died and 33.5 patients who lived would not have been treated according to the protocol. The absolute values of the differences between the observed numbers and these expected numbers all are 0.5; the calculated χ²-square statistic is 0.03 with 1 degree of freedom which is less than 3.84, the value associated with a 5% probability of the difference being due to chance alone. Therefore, Provit concluded that there was more than a 5% probability that the association of death and adherence to the antibiotic protocol was due to chance alone (ie, that the association between death and use of the protocol was not statistically significant).

The use of asymmetrical stratified sampling illustrated in the previous analyses is a useful technique for reducing the cost of data retrieval without sacrificing analytic power excessively. Because the rate of adverse outcomes is relatively low, all cases experiencing adverse outcomes are included in these analyses. On the other hand, limiting abstraction of medical records of cases without adverse outcomes to a random sample consisting of the same number of cases as the sample of cases with adverse outcomes reduced abstraction costs in this example by 80%. On the other hand, because changes in analytic power vary in proportion to the square root of the number of cases, elimination of most of a disproportionately large number of surviving cases had a relatively small effect on the results of these analyses.

Because of Dr. Lightouch’s refusal to accept results based on incomplete data, Provit bit the bullet, abstracted all 600 medical records, and repeated her analyses. Of the 540 patients who survived, 365 had received a general anesthetic and 175 had received a regional anesthetic (compared with estimates of 360 and 180, respectively, from the random sample). The absolute values of the differences between the observed numbers and these expected numbers all are 13.0; the calculated χ²-square statistic is 14.90 with 1 degree of freedom which is greater than 6.63, the value associated with a 1% probability of the difference being due to chance alone. This analysis confirmed Provit’s original conclusion that death was related to the use of a general anesthetic and that the probability that this association was due to chance alone was less than 1%.

Of the 540 patients who survived, 273 were treated according to the hospital’s antibiotic protocol and 267 were not (compared with estimates of 234 and 306, respectively, from the random sample). The absolute values of the differences between the observed numbers and these expected numbers all are 3.0; the calculated χ²-square statistic is 0.67 with 1 degree of freedom which is less than 3.84, the value associated with a 5% probability of the difference being due to chance alone. This analysis confirmed Provit’s original conclusion that there was more than a 5% probability that the association of death and adherence to the antibiotic protocol was due to chance alone (ie, that the association between death and use of the protocol was not statistically significant). In this case, a shift of almost 15% from the actual to the estimated number of cases treated according to the antibiotic protocol had a negligible effect on the computed χ²-statistic.

Simulated Illustrative Case Study: Part 2

Unfortunately for Dr. Lightouch, there is a fatal flaw in the preceding analyses. This flaw was detected by Amie Wisemann, an epidemiologist who was asked to review Provit’s analyses before she submitted her final report to the surgical peer review committee. Although Dr. Wisemann agreed that
the preceding analyses provided valid information about the relation between inpatient mortality and two clinical decisions, she raised serious questions about whether this information was sufficient to guide clinical practice at Cookbook Memorial. Instead, she recommended evaluating whether comparisons of care given to patients who died and care given to all or to a random sample of all patients discharged alive provides a fair basis for determining which interventions result in superior or clinical outcomes.

When Provot asked Dr. Wisemann to explain her reservations further, Dr. Wisemann indicated that in randomized control clinical trials, patients are randomized prior to an intervention in order to insure the comparability of treatment and control groups. If postintervention analyses reveal substantial, statistically significant differences in pretreatment characteristics between these groups, conclusions drawn from the study may be called into question. On the other hand, in observational comparisons of clinical interventions, interventions cannot be assumed to be randomly administered. Therefore, control cases must be selected carefully so that they have the same likelihood of being treated as do cases that actually were treated. Propensity analyses that match treated and untreated patients based on important descriptive characteristics are an excellent method of mimicking a randomized control clinical trial using observational data.  

By analogy, to obtain valid associations between outcomes and processes of care, analysts must ensure that the pretreatment probability of the outcome of interest is the same in patients who did and patients who did not experience that outcome. This requires that preintervention risk profiles of patients with and without adverse outcomes be compared. If these profiles differ substantially, then preintervention risks of patients who did and patients who did not experience an adverse outcome must be matched in order to create a valid comparative standard by which to judge the association between outcome and interventions (Figure 1).

To do this, Dr. Wisemann applied a risk-adjustment system developed by Provot. By analogy, to obtain valid associations between outcomes and processes of care, analysts must ensure that the pretreatment probability of the outcome of interest is the same in patients who did and patients who did not experience that outcome. This requires that preintervention risk profiles of patients with and without adverse outcomes be compared. If these profiles differ substantially, then preintervention risks of patients who did and patients who did not experience an adverse outcome must be matched in order to create a valid comparative standard by which to judge the association between outcome and interventions (Figure 1).

To correct for this systematic bias against general anesthesia and adherence to antibiotic protocols, Dr. Wisemann created a control group of survivors by matching each of the 60 patients who died to a patient who had a similar preoperative predicted mortality rate of 5.6% for patients receiving a regional anesthetic. Similarly, the average preoperative predicted mortality rate for patients receiving a general anesthetic was 11.9% compared to an average preoperative predicted mortality rate of 5.6% for patients receiving a regional anesthetic. Similarly, the average preoperative predicted mortality rate for patients who received prophylactic antibiotics according to protocol was 13.5% compared to an average preoperative predicted mortality rate of 6.5% for patients who did not receive prophylactic antibiotics according to protocol. In sum, the average preoperative predicted mortality rate for patients who received prophylactic antibiotics according to protocol was 13.5% compared to an average preoperative predicted mortality rate of 6.5% for patients who did not receive prophylactic antibiotics according to protocol.

To correct for the systematic bias against general anesthesia and adherence to antibiotic protocols, Dr. Wisemann created a control group of survivors by matching each of the 60 patients who died to a patient who had a similar preoperative predicted mortality rate of 5.6% for patients receiving a regional anesthetic. Similarly, the average preoperative predicted mortality rate for patients receiving a general anesthetic was 11.9% compared to an average preoperative predicted mortality rate of 5.6% for patients receiving a regional anesthetic. Similarly, the average preoperative predicted mortality rate for patients who received prophylactic antibiotics according to protocol was 13.5% compared to an average preoperative predicted mortality rate of 6.5% for patients who did not receive prophylactic antibiotics according to protocol.
53.5 patients who lived would have been given a general anesthetic, and 6.5 patients who died and 6.5 patients who lived would have been given a regional anesthetic. The absolute values of the differences between the observed numbers and these expected numbers all are 1.5; the calculated $\chi^2$-square statistic is 0.78 with 1 degree of freedom which is less than 3.84, the value associated with a 5% probability of the difference being due to chance alone. Therefore, Dr. Wisemann concluded that there was more than a five percent probability that the association between death and the type of anesthetic administered was due to chance alone (ie, that the association between death and the type of anesthetic administered was not statistically significant).

Similarly, when the 60 patients who died were compared to this matched sample of 60 patients who survived, Dr. Wisemann found that 27 of the patients who had died had been treated according to the antibiotic protocol and 33 had not, compared with 43 controls who had been treated according to the antibiotic protocol and 17 who had not. If adherence to the antibiotic protocol had no effect on survival, 35.0 patients who lived would have been treated according to the antibiotic protocol, and 25.0 patients who died and 25.0 patients who lived would have been treated according to the protocol. The absolute values of the differences between the observed numbers and these expected numbers all are 8.0; the calculated $\chi^2$-square statistic is 8.78 with 1 degree of freedom which is greater than 6.63, the value associated with a 1% probability of the difference being due to chance alone. Therefore, Dr. Wisemann concluded that death was related to failure to adhere to the hospital's antibiotic protocol and that the probability that this association was due to chance alone was less than 1%.

**Conclusion**

How could reasonable clinical impressions validated by Provit’s careful case-control analyses turn out to be so wrong? (In this case, the validity of Dr. Wisemann’s conclusions can be accepted without reservation because her findings conform to the operational specifications that were programmed into the simulator that produced the data presented in the scenario.) And what lessons does this scenario teach about how best to conduct surgical peer review?

First, drawing valid conclusions from in-depth analyses of single cases is fraught with difficulty. Instead of concentrating on detailed case studies, it is best to compare sets of similar cases with adverse outcomes to matched sets of cases without adverse outcomes. Second, matching should be based on pre-operative predicted adverse outcome rates. As described elsewhere in these proceedings, these predicted rates can be derived using risk-adjustment equations that require readily available enhanced administrative data that can be obtained electronically without expensive abstraction of medical records. Third, the size of control samples should be limited to reduce the cost of data acquisition. Fourth, peer review committees should devote their energy to formulating hypotheses about which processes might be critical in determining success or failure and helping to create a data collection protocol to obtain information about those processes. Finally, chart review should be performed not to aid in reviewing single cases or to risk-adjust clinical outcomes, but to link processes to outcomes as illustrated in the preceding scenario.

By focusing on sets of potentially important processes, critical practices can be identified and targeted for quality improvement initiatives. By relating proposed changes in policies and procedures to proven associations between clinical processes and patient outcomes, peer review committees can give scientific legitimacy to quality improvement efforts that might otherwise be viewed as more disruptive than potentially beneficial.

**References**


