Measuring and Improving the Quality of Care for Abdominal Aortic Aneurysm Surgery

Justin B. Dimick, MD, MPH; Gilbert R. Upchurch, Jr, MD

Because of widespread recognition of variations in surgical performance, patients and payers are placing renewed focus on holding providers accountable for the quality of care they provide. Patients are becoming more cautious informed consumers. Payers are stepping up efforts to align reimbursement to physicians and hospitals with the quality of care they provide. Providers can count on quality measurement and improvement being part of the health policy landscape for the foreseeable future. As stakeholders in this movement, we need to continually refine and update our understanding of the best approach to answer this call for accountability, and each of us needs to fulfill our obligation to contribute to the improvement of our clinical specialty.

The purpose of this article is to update our knowledge on the best methods to approach quality measurement and improvement in aortic surgery. Whereas many agree that improving the quality of care is important, few agree on the best way to go about it. The available approaches to measure the quality of care for high-risk abdominal aortic aneurysm surgery will be reviewed. In addition, the promise and pitfalls of existing improvement efforts will be considered.

Abdominal aortic aneurysm repair is both a relatively common and high-risk procedure. Many patients die each year after elective repair of abdominal aortic aneurysms. Improving the quality of care provided to these patients could potentially avert many of these deaths. For this reason, operations for aortic aneurysm disease are often the focus of quality measurement and improvement. Although most existing quality measurement initiatives focus on a narrow group of operations, they invariably include surgery for abdominal aortic aneurysms.

Measuring Quality

Before improving the quality of care in patients with aortic disease, it must first be decided on how to best measure it. Unfortunately, no consensus on how to go about measuring quality exists. A natural starting place is to introduce the Donabedian paradigm of structure, process, and outcomes. Structural elements include those attributes of the system in which the patient receives care. Process of care includes those specific details of how care was delivered to the patient. Outcomes are obviously the end results of the care provided.

Although this paradigm has stood the test of time, a fourth option will be discussed—combining individual measures to create a composite measure. Composite measures are some combination of structure, process, and outcomes. Using composite measures has advantages that make this approach uniquely suited for measuring surgical quality, and these will be outlined.

Structure

Structural elements that affect quality include attributes of the hospital or surgeon. Most previous studies focus on the relationship of hospital volume to operative mortality rates. Table 1 shows a summary of studies documenting the relationship of hospital volume to mortality rates for several vascular surgical procedures. Higher-volume hospitals consistently perform better than lower-volume hospitals for abdominal aortic surgery, with a 2-fold higher relative risk of mortality at low-volume hospitals (Table 1). No single threshold to define a high volume exists. Instead, it seems that the more you do, the better you get, with a stepwise decline in mortality with increasing volume. In most studies for abdominal aortic surgery, the highest volume group performs >30 to 40 cases per year. The median value following elective abdominal aortic aneurysm repair in a high-volume hospital for the 8 studies reviewed by Halm et al was 36 cases per year (Table 1).

Whereas most previous studies focus on open repair, an emerging body of evidence supports a similar relationship for endovascular aneurysm repair.7,8 This evidence is important, because the endovascular approach takes up a larger share of the market each year. We recently performed an updated analysis of volume and mortality using the Nationwide Inpatient Sample (NIS) from 2001 to 2003. We found a strong inverse relationship between hospital volume and adjusted mortality rates for both open and endovascular repair (P<0.001 for all comparisons). Hospitals in the lowest quintile of endovascular volume had mortality rates 3-fold higher than those in the highest quintile (2.4% versus 0.8%). Figure 1 shows the relationship of hospital volume and mortality in this updated analysis for all repairs together (labeled “overall”) and for both open and endovascular repair alone.
There also appears to be an inverse relationship between hospital volume and mortality rates with ruptured aneurysms when fixed via an endovascular approach. Greco et al used hospital discharge data sets from 4 states (Florida, Pennsylvania, New York, and New Jersey) to study the outcomes of ruptured aneurysms treated by an endovascular approach. The authors found much lower mortality rates in high-volume hospitals compared with low-volume hospitals (26% versus 46%).

A similar association has been found between individual surgeon volume and operative mortality. Many have questioned whether it is the surgeon’s or the hospital’s volume that is most important. A recent study provides the answer: both are equally important for abdominal aortic surgery. In an analysis of the national Medicare population, hospital and surgeon volume each explained approximately 50% of the variation in mortality rates across hospitals. This finding makes intuitive sense, because the surgeon lends his or her expertise and technical skill and the hospital system provides collective knowledge that impacts postoperative care (nurses, intensive care units, and diagnostic testing). The hospital system helps to avoid unnecessary deaths by preventing and rescuing patients from postoperative complications.

Surgeon training, specifically added qualification in general vascular surgery, independent of volume, is also associated with improved outcomes after abdominal aortic aneurysm repair. Using the Florida Agency for Health Care Administration state admission database, Pearce et al identified

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Studies, n (Statistically Significant)</th>
<th>Median Cases Defining Low Volume, n (Range)</th>
<th>Median Cases Defining High Volume, n (Range)</th>
<th>Mortality Rate, Median Average (Range)</th>
<th>Absolute Difference in Mortality for High vs Low Volume, Median (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carotid endarterectomy</td>
<td>15 (7)</td>
<td>10 (5 to 50)</td>
<td>50 (21 to 100)</td>
<td>1.8 (0.9 to 2.3)</td>
<td>0.4 (−0.05 to 1.8)</td>
</tr>
<tr>
<td>Elective abdominal aortic aneurysm repair</td>
<td>8 (7)</td>
<td>12 (3 to 31)</td>
<td>36 (12 to 433)</td>
<td>7.5 (3.8 to 7.6)</td>
<td>3.3 (1.1 to 11.6)</td>
</tr>
<tr>
<td>Ruptured abdominal aortic aneurysm repair</td>
<td>8 (2)</td>
<td>9 (2 to 10)</td>
<td>20 (5 to 50)</td>
<td>49.8 (40.0 to 63.2)</td>
<td>7.9 (1.5 to 18.7)</td>
</tr>
<tr>
<td>Lower extremity arterial bypass</td>
<td>2 (1)</td>
<td>13 and 20</td>
<td>32 and 100</td>
<td>3.1 and 3.8</td>
<td>1.1 and 1.4</td>
</tr>
</tbody>
</table>

*Modified from Halm et al.*

Figure 1. Relationship of hospital volume to operative mortality for all types of repair considered together and each type separately. Data are from the Nationwide Inpatient Sample and Dimick et al.
Table 2. Approaches to Measuring the Quality of Care for Aortic Surgery With Advantages and Disadvantages of Each Approach

<table>
<thead>
<tr>
<th>Example</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td>Hospital or surgeon volume</td>
<td>Inexpensive and readily available</td>
</tr>
<tr>
<td></td>
<td>Surgeon specialty training</td>
<td>Good proxy for outcomes</td>
</tr>
<tr>
<td>Process</td>
<td>Perioperative β-blockade</td>
<td>Actionable as targets for improvement</td>
</tr>
<tr>
<td></td>
<td>Endovascular approach to repair</td>
<td>Less influenced by patient risk and random errors</td>
</tr>
<tr>
<td>Outcome</td>
<td>Risk-adjusted mortality and morbidity rates</td>
<td>Seen as the bottom line of patient care</td>
</tr>
<tr>
<td>Composite</td>
<td>Need for reintervention after repair</td>
<td>Buy in from providers</td>
</tr>
<tr>
<td></td>
<td>Combination of multiple individual measures</td>
<td>Overcomes problems with small sample size</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Makes optimal use of all available information</td>
</tr>
</tbody>
</table>

In addition, increasing surgical specialization in vascular surgery, independent of board certification, was found to be associated with outcome for abdominal aortic aneurysm surgery. This study ascertained surgeon specialty using a profile of each surgeon’s practice (>75% vascular cases defined the specialty of vascular surgery). Using a nationally representative data set with 3912 patients undergoing abdominal aortic aneurysm repair in 1997, the authors found that lower mortality rates were achieved by vascular surgeons (2.2%) compared with cardiac surgeons (4.0%) and general surgeons (5.5%). After adjusting for hospital and surgeon volume, surgery performed by general surgeons was associated with a 76%-greater risk for death (95% confidence interval, 10% to 190%; \( P=0.009 \)).

Although most people would agree that these structural variables are important predictors of outcome, few agree on how to use this information to improve quality. The relative advantages and disadvantages of each measure need to be carefully considered. Table 2 shows the advantages and disadvantages of each type of measure included in this section. The primary advantage of using structural measures, such as hospital and surgeon volume, is that they are inexpensive and easy to obtain. Often they can be abstracted from readily available administrative data. Another advantage is that these structural measures are strongly correlated with important outcomes, including mortality and morbidity.

The greatest disadvantage of these measures is that they are not good at discriminating between individual providers—they are only good at detecting differences between groups of providers. This results in misclassification, with some low-volume hospitals providing high-quality care and some high-volume hospitals providing low-quality care. Another disadvantage is that volume and other proxy variables are not actionable from the provider perspective. As discussed later, these measures lend themselves more easily to selective referral strategies.

Another disadvantage of using volume as a quality indicator is the nature of the evidence base. Many studies are criticized because they are conducted using administrative data. When compared with prospective clinical data registries, administrative data sources have less clinical detail available for risk adjustment. As a result, it is often argued that differences in mortality rates between high- and low-volume hospitals may be due to differences in patient characteristics. Two reasons explain why this scenario is unlikely. First, many studies using clinical data have also shown a significant effect of volume on outcome. Second, the “null” volume–outcome studies show no effect of volume on outcome even without risk adjustment. Perhaps the most often cited study in this category is from the Veterans Affairs (VA) National Surgical Quality Improvement Program. If the lack of a volume–outcome effect were due to more detailed clinical risk adjustment, a difference would be found in the unadjusted outcomes between high- and low-volume hospitals—but a difference is not found. The lack of an effect of volume on outcome is therefore not due to better risk adjustment. Rather, it is likely because VA hospitals are all low volume, are staffed by surgeons who often operate across the street in a high-volume university setting, or share other similar structural elements or processes of care. In summary, it is possible that the estimates of volume–outcome effects found in studies based on claims data would be different using more detailed risk adjustment, but no data support the notion that the effects will disappear entirely.

**Process**

A dearth exists of process measures for abdominal aortic surgery. Indeed, few process measures are available for most surgical procedures. This situation stands in contrast to that for most medical conditions. For example, numerous process measures are available for acute myocardial infarction, pneumonia, and congestive heart failure. Most process measures relevant to surgery focus on elements of perioperative care common to all surgical procedures. For example, the Center for Medicare and Medicaid Service’s (CMS) Surgical Care Improvement Project (SCIP) focuses on appropriate use of antibiotic prophylaxis to prevent surgical site infection; the appropriate use of measures to prevent venous thromboembolism; and other processes related to the prevention of cardiac events and nosocomial pneumonia.

The one process measure specific to aortic aneurysm surgery is the proportion of appropriate patients who receive β-blockers in the perioperative period. However, the evi-
dence for this process measure is evolving. Although patients undergoing high-risk noncardiac surgery have been shown to benefit from the administration of β-blockade, newer data reveal that only patients with a history of myocardial ischemia are likely to benefit.17 Process measures will need to be continually refined to make sure they are consistent with the best available evidence.

Another process measure in abdominal aortic surgery is the type of approach used, open versus endovascular (Table 2). Using this as a quality measure is controversial. It is established that endovascular aortic aneurysm repair results in a lower short-term mortality rate. However, the endovascular method is less durable and has a much higher reintervention rate. In addition, the mortality difference appears to equal out after 2 years.18,19 As the body of data supporting the superiority of the endovascular approach for certain populations matures, this process measure will likely be touted as an important measure of quality. If the approach to repair is used as a quality measure, it will be important to ensure that the denominator is defined appropriately. Only a subset of patients is eligible for endovascular repair, and the denominator should be restricted to this group of patients. Because this level of clinical detail is not available from claims data, using this as a quality measure will require detailed clinical data to make sure the denominator is correct.

One process measure related to endovascular surgery that may prove useful is the proportion of patients who receive follow-up imaging. Unlike patients with open repair, all patients who receive endovascular repair should be followed up for with serial imaging for the duration of their lives to ensure secure repair. Consensus exists on this guideline, and such information would be readily available from some administrative data sources (eg, Medicare or insurance plan data sets).

Process measures have many advantages over other approaches at quality measurement. First, they are very actionable from the provider perspective. Low rates of adherence to process measures can easily be targeted for improvement. Second, they represent the care that was actually given to the patient and are viewed as “fair” by providers. They are less influenced by patient severity of illness and random chance. However, they also have several drawbacks. Unfortunately, we do not know the processes that contribute to optimal care for most operations, including abdominal aortic surgery. Most of the details of care that contribute to optimal care are not measurable using current methods.

Outcomes
Direct ascertainment of quality using risk-adjusted outcomes is widely perceived to be the gold standard of quality measurement. For high-risk surgical procedures, including abdominal aortic surgery, risk-adjusted morbidity and mortality rates are used to measure quality. Whereas this approach is attractive at first glance, several problems with direct outcomes measurement conspire to threaten the validity of this approach. These problems are all derived from a single root cause: poor quality is not the only driver of bad outcomes. Outcomes are a function of patient characteristics (risk factors), quality (hospital and surgeon factors), and chance (random variation). No one knows how much each of these factors contributes to observed variation in outcomes. The only solution is to control or adjust for these other factors as much as possible.

Several strategies of risk adjustment have been developed to take differences in patient factors into account. Clearly, we must measure and adjust for these important variables. Prospective ascertainment of patient risk factors using a clinical registry is considered the best method to collect data for risk adjustment. A logistic regression model is then used to calculate the expected number of deaths for each hospital. The ratio of observed to expected deaths can then be used as a risk-adjusted quality metric. However, the collection of detailed clinical data from the medical chart is expensive and prohibitive in many settings.

It is much more difficult to fully account for random variation when using outcomes to measure quality. This problem can be explained by the small caseloads (ie, small sample sizes) in individual hospitals: the smaller the hospital, the less likely extreme mortality rates are to truly represent good or bad quality. When a small hospital has no deaths, this is much more likely to be due to the small numbers of cases than perfect care.20 Traditionally, this problem is dealt with by using probability values or 95% confidence intervals. Most hospital report cards define poor quality as those hospitals with mortality rates that are significantly higher than average (or whose 95% confidence intervals do not overlap with the average).

It readily becomes apparent that this traditional solution only solves part of the problem when you consider the smallest hospitals. They will have the largest confidence intervals, and it will be very difficult to demonstrate a significant difference between these hospitals and a benchmark. Although the problem of small sample size has received considerable attention in the context of clinical trials, it has received less consideration in quality measurement. We performed a recent study to explore this problem for the 7 surgical procedures suggested for mortality measurement by the Agency for Healthcare Research and Quality.21 Using data from the Nationwide Inpatient Sample, we defined the national average mortality rate for each procedure and the number of cases performed in each hospital. The minimum sample size needed was entered to identify a poorly performing hospital as significantly different from the national average mortality rate. Finally, the proportion of US hospitals that exceed this minimum caseload (ie, those hospitals for which mortality would reliably reflect quality) was determined. The findings suggest that only coronary bypass surgery fulfills these criteria.21 For the other 6 operations in the analysis, including abdominal aortic aneurysm repair, less than half of hospitals in the United States perform enough cases to detect a doubling of the mortality rate (Figure 2).

Composite Measures
One approach for dealing with small samples is to combine multiple quality measures into a composite score. This approach also has the advantage of dealing with the problem of multiple measures and simplifying quality measurement by providing a single summary measure of performance. This
approach has several other advantages and some disadvantages (Table 2).

Composite scores can be created in 2 ways: (1) by combining information across procedures to provide a measure of aggregate performance, and (2) by combining information across measures to reflect performance with a single procedure. The approach of combining multiple operations to create an aggregate measure of performance is used by the National Surgical Quality Improvement Program (NSQIP).22 In the NSQIP, all operations are combined to create an overall risk-adjusted mortality and morbidity rate. Aggregating operations in this way increases the sample size and result in a much more precise measure with narrow confidence intervals. Unfortunately, it is hard to interpret the results. If a hospital is labeled an outlier for “overall” mortality, few clues are provided about where to find the quality problem. In addition, hospitals with poor quality for certain operations can be masked by their average performance with more common operations. Such hospitals will not be labeled as outliers and will miss out on the opportunity to improve.

The approach of combining multiple measures for a single operation will avoid this problem. This type of measure can take several forms. Perhaps the most simple is to simply sum the adverse events or nonadherence to process measures in an “all or none” format.23,24 This approach is currently used in many pay-for-performance initiatives.25 An alternative approach relies on a combination of domains based on empirically derived weights. The measures that most reliably predict subsequent mortality are afforded the most weight. Inputs for such a composite measure could include risk-adjusted mortality, risk-adjusted morbidity, hospital volume, surgeon volume, and perhaps performance with other related procedures. A general approach to creating these measures using empirical Bayes techniques has proven useful in several clinical contexts.26,27

Composite measures are currently used and will be more widely used in many upcoming policy initiatives. Most existing pay-for-performance efforts already use an all-or-none type of measure. The Society for Thoracic Surgeons Measurement Task Force is piloting a new composite score that combines elements of outcomes and processes of care into a single measure.3 The Agency for Healthcare Research and Quality has convened a working group to work on composite measures and recently has proposed a technique based on an empirically derived weighting system.28

**Figure 2.** Big problems with small samples: The proportion of hospitals in the United States with sufficient caseloads (sample size) to reliably use mortality rates to measure quality.

**Improving Quality**

Two basic options are available for improving the quality of care for patients undergoing abdominal aortic surgery. The first option is to send patients to hospitals and surgeons who provide the highest-quality care. The second option is to improve care at those hospitals where patients already receive their care. Neither solution is perfect. The potential benefits and downsides associated with each of the approaches will be discussed. Table 3 is a summary of the 2 approaches, examples of each, and their primary advantages and disadvantages.

**Selective Referral**

The chief advantage of selectively referring patients to the best hospitals and surgeons is that it is easy to envision how to do it. Most efforts at selective referral rely on proxy measures of quality, such as hospital or surgeon volume. As discussed in the previous section, these measures are easily determined from readily available administrative data. Selective referral can be carried out by simply providing information to patients and letting them make informed decisions. Alternatively, payers can create incentives or restrict access to steer patients to specific centers of excellence.

The disadvantage of selective referral is that many patients may not have anywhere to go. Limiting access to the highest-volume providers is problematic if you happen to be located in an area that does not have any. Some large regions of the United States do not have a single high-volume
Table 3. Approaches to Improving the Quality of Care for Aortic Surgery With Advantages and Disadvantages of Each Approach

<table>
<thead>
<tr>
<th>Example</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selective referral</td>
<td>Leapfrog Group: Advocates referral of patients to high volume hospitals</td>
<td>Proxy measures can be used</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No need to understand the complex mechanisms underlying good performance</td>
</tr>
<tr>
<td>Quality improvement</td>
<td>National Surgical Quality Improvement Program: Reports risk-adjusted mortality and morbidity rates</td>
<td>Buy in from surgeons</td>
</tr>
<tr>
<td></td>
<td>All hospitals can improve their performance</td>
<td>Difficult to measure procedure-specific outcomes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Many processes may not be portable across settings</td>
</tr>
</tbody>
</table>

Most of the experience with regional collaborative quality improvement comes from cardiac surgery. One of the most successful examples is the Northern New England Group, a coalition of hospitals performing cardiac surgery in this region. Through rigorous data collection and tracking of risk-adjusted mortality rates, feedback to hospitals, and sharing of best practices, this group has demonstrated reductions in the mortality rates since the inception of the program. The mechanism for improvement appears to be: (1) the systematic study of the relationship of processes of care to outcomes and then (2) the dissemination of these processes across all the hospitals.

More recently, the principles of continuous quality improvement have been used in noncardiac surgery. The most high-profile example is the NSQIP. The NSQIP was formed in VA hospitals with the goal of providing a platform for quality improvement. The NSQIP tracks risk-adjusted mortality and morbidity rates, provides feedback to hospitals, and encourages these sites to participate in quality-improvement efforts. Since the inception of this quality-measurement initiative, significant reductions in morbidity and mortality have been seen in VA hospitals.

However, no detailed accounting exists of how these improvements came about. The NSQIP is a good example of a system that expends a great deal of effort on measurement but has paid much less attention to the improvement part of the cycle. At least part of the difficulty with generating improvement is the nature of the quality measure used. As mentioned before, the NSQIP reports an aggregate mortality rate, which both acts to mask problems by diluting out the mortality experience for individual operations and also makes it difficult to know where to look for bad quality if you are flagged as a high-mortality outlier. As in the NSQIP, numerous examples exist of clinical registries within surgery and other disciplines that need to focus more on the improvement part of the cycle.

Whereas the NSQIP is a broad effort aimed at all surgical specialties, new registries have been specifically designed for improving quality in vascular surgery. For example, The Vascular Study Group of Northern New England (VSGNNE) consists of 45 vascular surgeons from 8 hospitals in Maine, New Hampshire, and Vermont. The VSGNNE has prospec-
Choosing the Best Approach

Whether selective referral or collaborative process improvement is the best approach to improving care will ultimately depend on how much quality is determined by measurable processes of care that can be transported across settings. If most existing variation in outcomes is due to processes that cannot be measured or are not portable, then very little improvement will result from collaborative quality improvement. Consider the 2 paradigms in Figure 3. In paradigm I, most existing variation in quality is due to measurable processes of care. In this paradigm, improving quality is only a matter of discovering those processes and ensuring they are disseminated to all hospitals. As a result, all hospitals/surgeons can provide optimal care for patients undergoing surgery. Now consider paradigm II, where most variation is due to unmeasurable processes of care or processes that are not portable across settings. These processes likely include the skill of the surgeon and the collective skill of all those who support the surgeon in their hospital system. In paradigm II, the measurable processes of care make up a much smaller piece of the pie. If this is the case, process improvement can only get us so far. It will certainly not get us to the point where all patients receive optimal care.

Ideally, we hope that we find ourselves in paradigm I, improving care at all centers through process improvement that is egalitarian and politically expedient. Unfortunately, evidence supporting paradigm II is accumulating. Most evidence comes from the literature of common medical conditions, such as acute myocardial infarction, for which more processes of care are known than any other medical condition. A recent study shows that only 6% of the observed variation in hospital mortality rates can be explained by the process measures used in the Joint Commission on Accreditation of Healthcare Organizations and Center for Medicare and Medicaid Services (JCAHO/CMS) set of measures.35

As more evidence accumulates, and some becomes available for surgical conditions, we may find observed differences in performance between good and bad hospitals only partly explained by known processes of care (paradigm II). Although it is important to continue to discover the processes of care that lead to higher quality, we should also be preparing to accept selective referral as a viable option in certain contexts. We owe it to our patients to strive for optimal care for everyone, whether it is delivered to them through quality improvement or they are delivered to it via selective referral. Most likely, however, both approaches will be needed as we strive to provide optimal care for surgical patients.

Disclosures

None.

References


Figure 3. Two paradigms for explaining the variation in performance across hospitals and surgeons are shown. In paradigm I, most existing variation is due to discrete, exportable, and measurable processes of care. In paradigm II, very little variation is due to measurable processes of care and most is due to surgeon skill and the collective knowledge of the hospital system, which are largely unmeasurable.


Key Words: aneurysm • aorta • outcome assessment • process measures • quality of health care • surgery