Does Volume Matter? The Effect of Trauma Surgeons’ Caseload on Mortality

Jack Sava, MD, Susan Kennedy, RN, Marion Jordan, MD, and Dennis Wang, MD

Background: Evidence suggests that trauma centers treating high volumes of severely injured patients produce lower mortality rates than those with low volumes. However, the effect of individual surgeons’ trauma caseload on outcomes has not been studied. This study compares outcomes between high-volume (HV) trauma surgeons admitting many patients with high injury severity, and low-volume (LV) surgeons treating fewer critical patients per year.

Methods: All trauma patients admitted to a large Level I trauma center over a 12-year period were assigned to either the HV or LV group, depending on the yearly volume of their admitting surgeon. Surgeons treating > 35 severely injured (Injury Severity Score > 15) patients per year were considered HV. Student’s t test and χ2 analysis were used to test comparability of LV and HV patient groups and to compare mortality rates. Mortality rates of HV and LV surgeons’ patients were compared in six injury patterns selected to represent moderate to severe injury. TRISS methodology (z score) was also used to assess outcomes in the two groups. The inherent bias of the TRISS method in comparing trauma outcomes was minimized by the homogeneity of the studied patient population.

Results: A total of 16,481 patients were admitted to HV surgeons, and 4,214 patients were admitted to LV surgeons. In all subgroups, HV and LV patients were similar regarding age, sex, physiologic status at admission, injury pattern, and injury severity. Mortality rates for HV and LV surgeons were not significantly different between the two groups in any injury pattern. The z score was 1.88 in the HV patient group versus 0.47 in the LV group.

Conclusion: Within a single institution, mortality rates for patients treated by surgeons admitting many severely injured patients were not significantly different from low-volume surgeons’ patients, although there was a trend toward higher mortality in the less active surgeons’ patients in some subgroups.

Key Words: Trauma, Caseload, Mortality, High volume, Low volume.


The injured patient fortunate enough to arrive at a Level I trauma center will find a number of important resources dedicated to his or her care, from organized triage to a dedicated trauma team and comprehensive rehabilitation. In all, 101 specific requirements for Level I certification are listed in the 1999 Resources for Optimal Care of the Injured Patient published by the Committee on Trauma (COT) of the American College of Surgeons (ACS). The sum of these criteria is state-of-the-art trauma care. However, some of the individual requirements are unproven, and many are controversial. Moreover, it is not known whether the improved care offered by certified trauma centers is the result of a few key ingredients, the additive effect of many subtle factors, or perhaps even a gestalt arising from the cumulative efforts of many interested professionals.

This article addresses one specific requirement for a trauma center: the minimum volume of patients with severe injury admitted yearly. The Committee on Trauma suggests a minimum of 35 severe cases per year (operative or nonoperative) for individual surgeons, or else an institutional total of 240 such cases. Within a single Level I trauma center, we compared the outcomes of surgeons admitting more than 35 patients yearly with an Injury Severity Score (ISS) greater than 15 with those admitting fewer than 35 such patients each year.

Patients and Methods

The trauma registry of the Washington Hospital Center was the source of all data. Washington Hospital Center is an urban Level I trauma center with a separate trauma receiving area; 24-hour in-house trauma attending coverage; surgical research programs; and training programs in surgical specialties, trauma, and critical care. Most years, penetrating trauma accounts for approximately 30% of admissions, and total admissions range from 1,500 to 2,500 patients per year. During the study period, between 13% and 25% (median, 20%) of admitted patients had an ISS > 15.

Daytime trauma coverage, in addition to postoperative care and the trauma educational curriculum, is provided by several full-time trauma surgeons. The night call roster includes several general or subspecialty surgeons who admit patients during their call shift, only to pass them to the full-time trauma attending staff the following morning. These are busy surgeons with active surgical practices, but they have varying involvement in trauma; some work only on
occasional weeknights, whereas others take frequent calls, including busy weekend nights.

To compare the outcomes of surgeons who met ACS individual surgeon volume criteria with those who did not, each surgeon was classified each year as either high-volume (HV) or low-volume (LV), on the basis of the number of severely injured (ISS > 15) patients they admitted. If in a given year a surgeon admitted more than 35 severely injured patients during that calendar year or the previous year, that surgeon-year was classified as HV. Thus, a surgeon who curtailed a busy trauma practice would be counted for a year afterward as HV, but thereafter would be counted as LV as long as his or her trauma caseload remained low. This adjustment was based on the assumption that any beneficial effect of frequent exposure to severe trauma is not immediately lost at the moment when a surgeon begins to take less trauma call.

All trauma patients were then split into two groups on the basis of the classification of their admitting surgeon, and their outcomes were compared. Although only exposure to severe trauma was used to categorize the surgeons as HV or LV, all trauma patients, regardless of injury severity, were used in the outcome comparison.

Two methods were used to compare outcomes. The TRISS methodology is the most commonly used technique for objectively analyzing trauma system performance. With this method, the patient’s age, anatomic injuries, and physiologic status on arrival are entered into a formula, along with a correction for penetrating versus blunt mechanism. By comparing this information to data derived from a large trauma database, a probability of survival is assigned to each patient. After analysis of the formula, the number of deaths predicted by TRISS can be compared with actual deaths to give a \( z \) score. The \( z \) score reflects the likelihood that actual deaths were significantly different from predicted. Typically, if \( z \) is between 1.96 and 1.96, it cannot be concluded that a real difference exists between actual and predicted deaths. When \( z \) exceeds 1.96, there are significantly fewer deaths than predicted by the model, and the converse is true when \( z \) is below 1.96. Using TriCode and Collector software (TriAnalytics, Bel Air, MD), the \( z \) score for patients admitted to HV surgeons was calculated and compared with the \( z \) for LV surgeons’ patients.

The second means of comparing outcomes between patients admitted to HV and LV surgeons was to arbitrarily choose a set of six injury patterns where early expert care might be expected to make a difference in survival. The injury patterns and registry search criteria were as follows: (1) gunshot wound to the abdomen with shock on arrival (gunshot wound abdomen and initial systolic blood pressure [SBP] < 90 mm Hg); (2) moderate to severe liver injury (Abbreviated Injury Scale [AIS] 1990 revision numerical injury identifier 541814.3, 541824.3, 541826.4, 541828.5, or 541830.6); (3) pelvic fracture with shock (pelvic AIS score > 1 and SBP < 90 mm Hg); (4) brain injury with coma (brain AIS > 1 and initial Glasgow Coma Scale score < 9); (5) thoracic injury requiring immediate surgery (thoracic AIS score > 1 and direct transfer from trauma bay to operating room for thoracotomy or median sternotomy); and (6) blunt abdominal trauma with shock (blunt mechanism and abdominal AIS score > 1 and initial SBP < 90 mm Hg).

Mortality for each injury pattern was then compared between patients treated by HV surgeons and those treated by LV surgeons, using \( \chi^2 \) analysis. To ensure comparability, patient groups were analyzed to verify that they had similar demographics, number of associated injuries, physiologic status on arrival, and overall injury severity. Mean values were compared using Student’s \( t \) test.

**RESULTS**

Between 1990 and 2001, 20,695 trauma patients were entered into the registry. Twenty-one surgeons took trauma night call during the 12-year period. The number of severely injured patients (ISS > 15) admitted by individual surgeons each year ranged from 1 to 104 patients per year (median, 31 patients per year). Average severe trauma admissions for each surgeon are shown in Figure 1. The average surgeon volume of severely injured patients (total patients with ISS > 15/total surgeons taking trauma call) ranged from 29 to 53 over the study period. In 7 of the years studied, the average surgeon volume of severely injured patients was less than 35, because of a few surgeons taking very infrequent call as substitutes, and admitting very few patients. Four surgeons admitted > 35 severely injured patients in some years, and fewer in other years. They were counted as LV when their yearly admissions were low and as HV when they met the yearly volume threshold. In all, 65 surgeon-year combinations were classified as HV and 51 as LV. When all patients were divided into two groups on the basis of the classification of their admitting surgeon in the year they were admitted, 16,481 patients were admitted to HV surgeons and 4,214 to LV surgeons.

As seen in Table 1, mortality was higher for LV surgeons’ patients in four of six groups compared, although it did not reach statistical significance. Demographic, anatomic,
and physiologic variables were not significantly different between any groups compared. The z score was 1.88 in the HV patient group versus 0.47 in the LV group.

**DISCUSSION**

Over the last 10 years, the surgical literature has given rise to a number of studies suggesting that some types of patients, diseases, and operations are best treated at high-volume “centers of excellence,” where expertise and resources are concentrated to give optimal, cost-effective care. This approach has been recommended for coronary bypass grafting, pancreatic resection, aortic aneurysm repair, transplantation, colon resection, and more than 20 other procedures in addition to the management of trauma.

Several articles have addressed the relationship between institutional volume and mortality among trauma centers. Smith et al. showed an inverse relationship between volume and mortality among Chicago’s trauma centers over 2 years. The study was later criticized for inadequate matching of case severity, the bane of all studies comparing interhospital outcomes. Waddell et al. used TRISS methodology to show that small Canadian trauma centers can match the performance standard of the Major Trauma Outcome Study database. However, the TRISS method has been widely criticized as a poor tool for trauma center comparison, with a bias toward smaller centers.

Konvolinka et al. used stepwise regression to identify relationships between mortality and various volume-related variables in Pennsylvania trauma centers. They found an association between mortality in a given trauma center and the average surgeon’s volume of seriously injured patients within that center. This association was most dramatic in bluntly injured patients. No distinction was made between high- and low-volume individual surgeons within institutions; however, the institutional volume in each center was divided by the number of surgeons to give each surgeon’s volume.

Tepas et al. reached a different conclusion. They compared the outcomes of children in the National Pediatric Trauma Registry to mortality rates predicted by the Pediatric Risk Indicator in low-, medium-, and high-volume trauma centers. They found that the medium-volume centers had the best outcomes, and hypothesized that extra deaths might occur in high-volume centers when resources are exhausted. Similarly, Cooper et al. analyzed New York State Trauma Registry Data and, after creating a unique model for predicting mortality within that data set, compared outcomes between centers. They found a trend toward better outcomes in lower volume centers. Neither of these studies addressed individual surgeons’ volume.

Rather than address the effect of institutional volume, the present study evaluates the impact of the admitting surgeon’s trauma activity on mortality. This is relevant for several reasons. First, we need to know what parts of a trauma system are most important, and one might assume that the experience of the surgeon present on the trauma patient’s arrival would be a key factor in the success of high-level trauma care. Also, many trauma centers are at least partially staffed by general surgeons taking trauma call, and the impact of their participation remains unstudied.

In dichotomizing surgeon caseload, we chose a cutoff of 35 cases per year with an ISS > 15. This is the standard set by the ACS-COT and is justified to a degree by the volume-per-surgeon analysis in the study by Konvolinka et al. We adjusted for surgeons who had recently cut back on trauma call but who had high trauma caseloads only a short time before and were presumably still benefiting from their previous experience.

Comparing outcomes between physicians is a thorny and contentious problem. We were able to sidestep many of the hazards of such comparisons because of the uniform case mix in our study. High-volume and low-volume patients were

### Table 1: Mortality in High- and Low-Volume Surgeons’ Patients, by Injury Pattern

<table>
<thead>
<tr>
<th>Injury Pattern</th>
<th>HV Mortality (%)</th>
<th>LV Mortality (%)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gunshot abdomen and shock on arrival*</td>
<td>62/142 (43.6)</td>
<td>24/45 (53.3)</td>
<td>0.09</td>
</tr>
<tr>
<td>Moderate/severe liver injury**</td>
<td>35/161 (21.7)</td>
<td>14/39 (35.9)</td>
<td>0.10</td>
</tr>
<tr>
<td>Pelvic fracture and shock on arrival***</td>
<td>33/62 (53.2)</td>
<td>15/24 (62.5)</td>
<td>0.20</td>
</tr>
<tr>
<td>Brain injury and coma†</td>
<td>503/1,134 (44.7)</td>
<td>114/269 (42.3)</td>
<td>0.60</td>
</tr>
<tr>
<td>Thoracic injury requiring emergency surgery†††</td>
<td>281/478 (58.8)</td>
<td>71/121 (58.6)</td>
<td>1.00</td>
</tr>
<tr>
<td>Blunt abdominal trauma + shock†††</td>
<td>48/249 (23.6)</td>
<td>16/46 (34.8)</td>
<td>0.17</td>
</tr>
</tbody>
</table>

GCS, Glasgow Coma Scale.

* HV and LV groups were not significantly different in age, gender, mean ISS, proportion with very severe injury (ISS > 25), mean GCS score, or proportion with extra-abdominal injury.

** HV and LV groups were not significantly different in age, gender, mean ISS, proportion with very severe injury (ISS > 25), mean GCS score, proportion with GCS score < 9, proportion with extra-abdominal injury, or proportion with liver AIS score > 4.

*** HV and LV groups were not significantly different in age, gender, mean ISS, proportion with very severe injury (ISS > 25), mean GCS score, proportion with GCS score < 9, proportion with extrapelvic injury.

† HV and LV groups were not significantly different in age, gender, mean ISS, proportion with very severe injury (ISS > 25), mean GCS score, proportion with GCS score < 9, proportion with extracranial injury, or proportion with shock on arrival.

†† HV and LV groups were not significantly different in age, gender, mean ISS, proportion with very severe injury (ISS > 25), mean GCS score, mean chest AIS score, or proportion with shock (SBP < 90) on arrival.

††† HV and LV groups were not significantly different in age, gender, mean ISS, proportion with very severe injury (ISS > 25), or mean GCS score.
admitted over the same time period to the same center, with the same prehospital care, nurses, ventilators, and resistant organisms. We were therefore able to compare crude mortality in a number of injury patterns with little risk of confounding. Likewise, the uniformity of case mix between groups made the TRISS analysis possible, avoiding the bias that this methodology brings to comparisons between trauma centers.

We hypothesized that patients admitted to high-volume trauma surgeons would have lower mortality than those admitted to low-volume surgeons. Our results did not support this conclusion. No significant improvement in mortality was seen in any patient group, although there were trends toward better outcomes in the high-volume group in all injury patterns except brain injury with coma and thoracic injury requiring immediate surgery. The TRISS analysis did not suggest that either high-volume or low-volume surgeons’ patients deviated significantly from outcomes predicted from the Major Trauma Outcome Study database.

Although we avoided the problem of heterogeneous case mix, there are several other limitations to this study. Surgeon activity was dichotomized, and differences between the extremes of trauma caseload might have been obscured by the use of only two categories. Also, the sample sizes of the groups in each injury pattern were small, and the mortality differences might have reached statistical significance in a study with more power. The TRISS analysis, too, might have revealed a significant difference in a larger data set. As always, failure to prove a difference does not constitute proof of sameness.

The study focuses on an arbitrary group of injury patterns. The real impact of admitting surgeons’ experience might be better shown in some other types of injury that we did not include. Also, the two patient groups in this study differed only with respect to the early hours of their care—all patients were on the same service starting the morning after admission. Finally, although the low-volume surgeons did not meet the minimum volume suggestions of the ACS-COT, most of the surgeons in this group were active in the trauma program and participated in morbidity and mortality and other weekly conferences. They therefore may not have been representative of the typical part-time trauma surgeon.

What are the necessary ingredients for optimal trauma care? The current literature does not tell us. Within the scope of its limitations, the present study suggests that a surgeon taking infrequent trauma call can provide initial care without a significant decrement in outcome, provided that he or she has the collective resources of a large Level I trauma center at his or her disposal. It should not be interpreted as evidence that good surgeons or trauma centers do not matter—all patients in this study had the benefit of both. Although it is perhaps not surprising that a single caregiver’s impact is overshadowed by the many other facets of a patient’s hospitalization, we nonetheless might expect that the leader of the team during such a critical time would impact outcome. This expectation is reflected in the current ACS-COT guidelines but was not supported by our data.

The ACS-COT recommendations have been made in light of previous work13,14 showing that centers where the average surgeon admits many severely injured patients have better results. However, ours is the only study where the surgeon’s volume was the only variable, and we were not able to show a statistically significant improvement in outcome with increased surgeon trauma caseload. Further study should assess the relative importance of the different facets of organized, multidisciplinary trauma systems, so that subsequent ACS-COT criteria can accurately define high-level care.

REFERENCES


DISCUSSION

Dr. Richard J. Mullins (Portland, Oregon): In their observational study, the authors demonstrated that the outcome, namely, hospital survival, of patients managed by high-volume surgeons and low-volume surgeons were equivalent. The implication the authors give this observation in the discussion is that the low-volume surgeons did as well as the high-volume surgeons. They should remain open to the alternative explanation, which is that the high-volume surgeons are not experts and do no better than surgeons who take trauma call on an occasional basis.

What do the authors think constitutes “expertise” in a high-volume surgeon? Dr. Sava, can you be specific as to what you think your data suggest are the skills of an expert trauma surgeon?

Given that severely injured patients often have multiple injuries, and thus multiple decision makers, why is it reasonable to expect that a single individual who admits the patient would have a dominant influence on survival? What was the volume of experience of the residents who were working with these faculty surgeons? Do you have a residency in which the most senior resident on trauma call with the faculty is in fact very experienced, and is the equivalent of a high-volume trauma surgeon? In other words, are the high-volume chief residents keeping the low-volume faculty out of trouble?

In comparing the high- and low-volume surgeons, should an alternative measure of performance to patient survival be evaluated? My bias is that high-volume surgeons are able to accomplish the same outcome without spending as much time or money on superfluous tests. Dr. Sava, do you think there is an efficiency value to having high-volume surgeons on trauma call? What is your conclusion from this review—should the ACS drop surgeon volume as a mandatory resource for optimal care of the injured patient?

Dr. Michael L. Hawkins (Augusta, Georgia): You didn’t tell us when they died, and was there a difference in the time of death between the low-volume and high-volume surgeons? Perhaps what you’re really measuring is the quality of your in-hospital care, and that your critical care team is able to overcome the deficits of the low-volume surgeon.

Dr. Krista L. Kaups (Fresno, California): Do you have any information about patients who survived to discharge, whether they were discharged home, to extended care, or to rehabilitation? As we know, for a head injury patient, the effect of appropriate resuscitation is ongoing and significant.

Dr. Michael A. West (Chicago, Illinois): I think this is a very important question, and I think that perhaps you’ve missed an opportunity to answer it. In terms of your question about high versus low, you had many in the middle that were only 2 or 3 cases below the “target” of 35 cases per year.

In your caveats, you hinted that maybe looking at the highest versus the lowest would answer the question. I wonder whether you did look at, say, the highest quartile versus the lowest quartile to see whether the “really experienced” versus the “really inexperienced” made a difference. If you still see no difference, I would go along with your conclusion, and it could be made more forcefully.

Dr. Matthew C. Indeck (Danville, Pennsylvania): I’m agreeing with a lot of people, that it’s tough to sort out the blunting effect of a resident’s team on the system’s approach to trauma care. The question comes down to this: can you measure a trauma surgeon’s ability to resuscitate a patient who has either a lot of operative experience or low operative experience?

Are the low-volume trauma surgeons who perform a lot of general surgery more skilled at the operative resuscitative issues as opposed to the high-volume trauma surgeons, who treat a lot of blunt trauma and have very low operative experience but are more skilled in the resuscitative phase? Were you able to sort out that difference?

Dr. Marc J. Shapiro (St. Louis, Missouri): Your article should probably be entitled “The Trauma Surgeon Meeting the Patient in the Emergency Room,” because you say you turn all these patients over to a trauma team, which would be overseen by high-volume trauma surgeons, the day after. How many of these patients required secondary operations for missed injuries or required other interventions from the high-volume trauma surgeons?

Dr. David B. Hoyt (San Diego, California): Did you correct your mortality per surgeon by the preventability-of-death analysis that you probably performed as part of your questions and answers?

Dr. Jack Sava (closing): Thank you, Dr. Mullins. First, with respect to what an “expert trauma surgeon” is, this may be changing. It now includes skills such as resource allocation, leading a multidisciplinary team, and managing the
patient’s course throughout the hospital stay. In the operating room, I would guess that the definition would include a healthy dose of common sense, mastery of some basic surgical maneuvers, and probably some skill at prioritization.

The question of the hour, though, is whether or not it matters if you have an expert trauma surgeon. On first glance, it seems an offensive question, and if you’re the guy who gets saved by a masterful atrio caval bypass, then the answer is that it matters quite a bit. However, if you look at it from a public health perspective or a funding perspective or a resource allocation perspective, it may be that the “number needed to treat” to see the effect of virtuosic technical surgery is pretty high.

With respect to multiple decision makers, there are often many—a lot of chefs stir the pot. These consultants probably have a substantial effect on cost, length of stay, functional outcome, and maybe length of stay. However, at least in the centers where I’ve worked, the onus has been on the trauma surgeon to make all decisions that could affect mortality, which was the endpoint in this study.

The residents at our center generally have a lot of trauma experience, with some variation depending on their rotations. Over the years, most residents end up working with all the attending physicians. It is possible that the skill of the residents—I don’t know if “compensate” is the right word—blunts the effect of the different attending physicians, as do the nurses, the waiting angiography team, the ready operating room, and the capable blood bank. In other words, the message I took home from this study is that the effect of the system is so large that it eclipses whatever effect the admitting surgeon may have, at least on overall morality.

Should we use a cost analysis? It would have been most worthwhile for all of us if we could have easily shown that volume and experience directly influence life and death. However, if we’re not able to do that, I guess we’ll have to settle for trying to show that our skill saves health care money, so that may be an appropriate direction for further study.

Finally, should we scrap the ACS volume requirement? I think that, as Dr. Maier suggested the other day, we don’t yet have a valid or widely accepted way of assessing the outcomes of a trauma center or a putative trauma center. Instead, we’re using surrogate markers of quality care, such as volume. As such, I don’t know that the individual surgeon volume requirement is any better or worse than the other 100 requirements listed in the Gold Book. Probably, until we have something better, that’s what we need to use. I doubt that there’s anyone claiming that the current ACS certification process is perfect and flawless at assessing trauma care, but I think that that means that we should critique it and improve it, not abandon it.

Dr. Hawkins, I think that I would plead the same answer to that as I did with the residents, that I think that’s probably true. I think that, like many other factors in the hospital, there’s a big safety net in a well-constructed trauma center that tends to eclipse the effect of an individual trauma surgeon. To be specific, I don’t know when they died. We used registry information. We didn’t have that.

Dr. Kaups, yes, as I mentioned, we looked only at death. I don’t have any information on the many other outcomes that would have been very interesting, such as functional outcomes, complications, or resource use.

Dr. West, it’s a question of picking a methodology. What I’ve found is that the nemesis of measuring outcomes in trauma is case mix variation. Everybody wants to find a fair way to compare one group of patients with another, but it’s very elusive. The unique quality of this method is that, because it only involves one center, we can pretty safely assume that the two groups of patients are quite similar, without confounding differences. This was borne out by statistical comparison. However, limiting the population to one center puts a very firm limit on the size of the groups and therefore the power of the study. I think it would have been very interesting to stratify the patients in a number of ways, including further subdivision of cases per year, but it would have sucked all the power out of the study.

Dr. Indeck, no. Actually, I was just thinking of that last night. It would have been interesting to know what happened to the patients who had operations. Again, though, if you teased through all the data to find truly comparable cases, I think there would be a sample size problem.

Dr. Shapiro, of course the subsequent care of a trauma patient is very important, and many of these patients do need further operations later in their stay. However, the specific focus here was whether or not a low-volume surgeon can take night call. All the patients had the same care thereafter, and reoperation was just one of many aspects of their continued care that was the same in all patients. That left the single variable of who was on call and admitted the patient.

Dr. Hoyt, the preventability is built in to the z score. Other than that, no, we did not.