The Trauma Quality Improvement Program: Pilot Study and Initial Demonstration of Feasibility

Mark R. Hemmila, MD, Avery B. Nathens, MD, PhD, Shahid Shafi, MD, MPH, J. Forrest Calland, MD, David E. Clark, MD, MPH, H. Gill Cryer, MD, PhD, Sandra Goble, MS, Christopher J. Hoeft, BS, J. Wayne Meredith, MD, Melanie L. Neal, MS, Michael D. Pasquale, MD, Michelle D. Pomphrey, RN, and John J. Fildes, MD

Objective: The American College of Surgeons Committee on Trauma has created a "Trauma Quality Improvement Program" (TQIP) that uses the existing infrastructure of Committee on Trauma programs. As the first step toward full implementation of TQIP, a pilot study was conducted in 23 American College of Surgeons verified or state designated Level I and II trauma centers. This study details the feasibility and acceptance of TQIP among the participating centers.

Methods: Data from the National Trauma Data Bank for patients admitted to pilot study hospitals during 2007 were used (15,801 patients). A multivariable logistic regression model was developed to estimate risk-adjusted mortality in aggregate and on three prespecified subgroups (1: blunt multisystem, 2: penetrating truncal, and 3: blunt single-system injury). Benchmark reports were developed with each center's risk adjusted mortality (expressed as an observed-to-expected [O/E] mortality ratio and 90% confidence interval [CI]) and crude complication rates available for comparison. Reports were deidentified with only the recipient having access to their performance relative to their peers. Feedback from individual centers regarding the utility of the reports was collected by survey.

Results: Overall crude mortality was 7.7% and in cohorts 1 to 3 was 16.4%, 12.4%, and 5.1%, respectively. In the aggregate risk-adjusted analysis, three trauma centers were low outliers (O/E and 90% CI <1) and two centers were high outliers (O/E and 90% CI >1) with the remaining 18 centers demonstrating average mortality. Challenges identified were in benchmarking mortality after penetrating injury due to small sample size and in the limited capture of complications. Ninety-two percent of survey respondents found the report clear and understandable, and 90% thought that the report was useful. Sixty-three percent of respondents will be taking action based on the report.

Conclusions: Using the National Trauma Data Bank infrastructure to provide risk-adjusted benchmarking of trauma center mortality is feasible and perceived as useful. There are differences in O/E ratios across similarly verified or designated centers. Substantial work is required to allow for morbidity benchmarking.

Key Words: Trauma outcomes, NTDB, TQIP, Quality improvement.

(*J Trauma.* 2010;68: 253–262)

The American College of Surgeons Committee on Trauma (ACS-COT) is committed to improving all phases of the management of the injured patient.[1] In 1979, the ACS-COT published the first edition of its resource document, entitled "Optimal Resources for the Care of the Injured" in which a systems approach to trauma care is emphasized and provided the framework for the trauma center verification review process.[2] At the heart of any effective trauma system is a multidisciplinary performance improvement (PI) process. The PI process is charged with monitoring patient care, evaluating adverse outcomes, and improving compliance with quality of care indicators. To support the PI process, the ACS-COT PI and Patient Safety committee has published a reference manual outlining the structure and criteria of an effective PI program.[3]

Conducted from 1982 to 1989, the Major Trauma Outcome Study (MTOS) established national standards for trauma care, which could be used to benchmark trauma centers.[4] The MTOS database allowed development of the Trauma Injury Severity Score (TRISS) methodology for predicting the probability of survival for an individual trauma patient. The prediction equation was based on indices of physiologic derangement, anatomic injury severity, and age. TRISS allows trauma centers to compare their observed outcomes to those predicted by the patient's presenting clinical status. However, advances in trauma and critical care have made the coefficients, calculated and validated with MTOS data, obsolete.

The National Surgical Quality Improvement Program (NSQIP), created in 1994, is the best known effort to improve

Submitted for publication September 24, 2009.
Accepted for publication December 15, 2009.

From the Department of Surgery (M.R.H.), University of Michigan Health System, Ann Arbor, Michigan; Department of Surgery (A.B.N.), University of Toronto, Toronto, Ontario, Canada; Department of Surgery (S.S.), Baylor Health Care System, Grapevine, Texas; Department of Surgery (J.F.C.), University of Virginia, Charlottesville, Virginia; Department of Surgery (D.E.C.), Maine Medical Center, Portland, Maine; Department of Surgery (H.G.C.), University of California Los Angeles Medical Center, Los Angeles, California; the American College of Surgeons (S.G., C.J.H., M.L.N.), Committee on Trauma, Chicago, Illinois; Department of Surgery (J.W.M.), Wake Forest University, Winston-Salem, North Carolina; Department of Surgery (M.D.P.), Lehigh Valley Hospital, Lehigh, Pennsylvania; Department of Surgery (M.D.P.), University of Virginia, Charlottesville, Virginia; and Department of Surgery (J.J.F.), University of Nevada, Las Vegas, Nevada.
This work was supported by National Institutes of Health grant K08-GM078610 with joint support from the American College of Surgeons and the American Association for the Surgery of Trauma (to M.R.H.) and by a Canada Research Chair in Systems of Trauma Care (to A.B.N.).
Presented at the 68th Annual Meeting of the American Association for the Surgery of Trauma, October 1–3, 2009, Pittsburgh, Pennsylvania.
Reprints: Mark R. Hemmila, MD, Trauma Burn Center, University of Michigan Health System, 1B407 University Hospital, 1500 E. Medical Center Drive, SPC 5033, Ann Arbor, MI 48109-5033; email: mhemmila@umich.edu.

DOI: 10.1097/TA.0b013e3181cfc8e6

the quality of surgical care using a risk-adjusted benchmarking methodology.[5,6] Since the inception of NSQIP, the 30-day postoperative mortality and morbidity in the Veterans Administration (VA) Health System has declined by 27% and 45%, respectively.[7] The initial success of NSQIP in the VA led to its expansion into non-VA private hospitals as part of the Patient Safety in Surgery (PSS) Study. The Patient Safety in Surgery study demonstrated a significant decline in morbidity of 8.7% during the 3-year period of the study.[8] Based on the early success of NSQIP, the ACS has assumed management of the program and expanded it to >200 hospitals throughout the United States.

In October 2006, the chairman of the ACS-COT, John Fildes, MD, FACS, convened an ad hoc work group to explore the concept of creating and implementing a risk-adjusted quality improvement program for trauma. The premise was that there were existing mechanisms and resources that could be drawn on such as the current trauma registry infrastructure, the National Trauma Data Bank (NTDB), and the National Trauma Data Standard (NTDS) to create a "Trauma Quality Improvement Program" (TQIP). Therefore, a practical goal of TQIP is to leverage the existing trauma registry and abstractor system along with the resources of the NTDB to create a benchmarking program on a national basis.

The TQIP work group was given a mandate to design, test, and implement a quality improvement program for trauma that is validated, risk-adjusted and outcomes based, to measure, and continually improve the quality of trauma care. To date, the TQIP advisory committee has produced executive reports for both the COT and ACS, provided a business plan to the ACS, embarked on a 3-year pilot study of feasibility, sponsored multiple trauma registrar training sessions, and is in the process of recruiting centers for participation in a national roll out. The data outlined in this article highlights results from the first year of the 3-year TQIP pilot study.

MATERIALS AND METHODS

The TQIP pilot study began in June 2008. A total of 23 trauma centers volunteered and were selected for participation (Table 1). To participate, centers were required to have undergone ACS or regional verification and subsequent Level I or Level II designation. Centers were selected based on their interest in TQIP and prior commitment to the NTDB. Trauma registrars and data abstractors from participating centers underwent training in June 2008. The registrar training course provided an overview of the objectives and infrastructure for TQIP. It also covered the data definitions and critical data fields for collection in each local trauma registry. Use of the NTDS data definitions was emphasized and data accuracy/completeness issues were addressed. Follow-up training has occurred using webinars, conference calls, and test case data abstraction.

Simultaneous with registrar training, 2007 calendar year data were obtained from the NTDB for centers participating in the study. These data were used for the analysis and are from the first year of this pilot study, which precedes TQIP registrar training. As such, the first year of data collected represents baseline results before implementation of focused TQIP registrar training or clear emphasis on the NTDS fields and definitions. These baseline results will allow future evaluation of the effectiveness of registrar training on data quality.

TABLE 1. Participating Trauma Centers

Names	Level
Cedars-Sinai Medical Center, Los Angeles, California	I
Christiana Hospital, Newark, Delaware	I
Genesys Regional Medical Center, Grand Blanc, Michigan	II
John Muir Medical Center, Walnut Creek, California	II
Lahey Clinic, Burlington, Massachusetts	II
Lehigh Valley Hospital, Allentown, Pennsylvania	I
Maine Medical Center, Portland, Maine	I
Massachusetts General Hospital, Boston, Massachusetts	I
Oklahoma University Medical Center, Oklahoma City, Oklahoma	I
Parkland Health and Hospital System, Dallas, Texas	I
Regional Medical Center at Memphis, Memphis, Tennessee	I
Ronald Reagan UCLA Medical Center, Los Angeles, California	I
Saint Mary's Health Care, Grand Rapids, Michigan	II
Sharp Memorial Hospital, San Diego, California	II
St. John Medical Center, Tulsa, Oklahoma	II
St. Michael's Hospital, Toronto, Ontario, Canada	I
St. Vincent Mercy Medical Center, Toledo, Ohio	I
Truman Medical Center, Kansas City, Missouri	I
University Medical Center, Las Vegas, Nevada	I
University of California, San Diego Medical Center, San Diego, California	I
University of Michigan, Ann Arbor, Michigan	I
University of Virginia, Charlottesville, Virginia	I
Wake Forest University Baptist Medical Center, Winston-Salem, North Carolina	I

The study cohort consisted of patients admitted to the participating trauma centers between January 1 and December 31, 2007. These data were submitted to the NTDB during the regular 2008 call for data. Inclusion and exclusion criteria for patients incorporated in the TQIP analyses are outlined in Table 2. The inclusion and exclusion criteria were selected to allow creation of a similar group of patients for analysis and eliminate bias caused through use of varying entry criteria for patients within each center's trauma registry. We also sought to exclude minimally injured patients from the analysis by requiring an injury severity score (ISS) of ≥9.

All ISS values were calculated by the NTDB using a validated crosswalk software program that derives abbreviated injury scale (AIS) codes from International Classification of Diseases, 9th Revision (ICD-9) diagnosis codes (ICDMAP-90, 1995 update, The Johns Hopkins University and Tri-Analytics, Baltimore, MD).[9] All centers submit ICD-9 injury diagnosis codes, but not all centers perform AIS coding and those that do use a number of different versions. Thus, ICD-9 to AIS mapping allows for consistent evaluation of ISS across centers.

TABLE 2. Inclusion and Exclusion Criteria

Inclusion criteria
Age ≥16 yr.
At least one valid trauma ICD-9-CM code in the range of 800–959.9 excluding late effects (905–909.9), superficial injuries (910–924.9), and foreign bodies (930–930.9).
Primary mechanism of injury classified as either blunt or penetrating
Blunt is defined as an injury where the primary E-code is mapped to the following categories: fall, machinery, motor vehicle traffic, pedestrian, cyclist, and struck by or against.
Penetrating is defined as an injury where the primary E-code is mapped to the following categories: cut/pierce and firearm.
Severely injured patients with at least one AIS ≥3
For blunt injuries: at least one injury in any of the following AIS body regions: head, face, neck, thorax, abdomen, spine, or upper/lower extremity.
For penetrating injuries: At least one AIS ≥3 injury in any of the following AIS body regions: neck, thorax, and abdomen.
Calculated injury severity score (ISS) ≥9.
ED discharge disposition and hospital discharge disposition must be known.
Exclusion criteria
GSW to the brain defined by: Any E-code: E922.0-922.9, E955.0-955.4, E965.0-965.4, E979.4, E985.0-985.4, E970, and at least one ICD-9-CM code in the range: 800–801.99 or 850–854.1.
Comorbidity: preexisting advanced DNR directive to withhold life sustaining interventions.
Dead on arrival.

GSW, gunshot wound; DNR, do not resuscitate.

TQIP analysis and reports were created for all the patients who met inclusion criteria. In addition to aggregated data, we also categorized patients into three distinct cohorts: (1) blunt multisystem injury (blunt mechanism with AIS ≥3 in at least two of the following AIS body regions: head, face, neck, thorax, abdomen, spine, and upper or lower extremities); (2) penetrating truncal injury (penetrating mechanism with injuries of AIS ≥3 in at least one of the following AIS body regions: neck, thorax, or abdomen); (3) blunt single-system injury (blunt mechanism with injuries of AIS ≥3 limited to only one AIS body region with all other body regions having a maximum AIS ≤2). Selection of these three separate cohorts was done to reflect the wide spectrum of trauma patients and their distinct challenges. Parsing the analysis into specific groups allows each center to assess its system performance from different process aspects (operative versus critical care management) and provides an opportunity for centers with significant overrepresentation of a particular type of patient to better understand their performance relative to peer centers.

The primary outcome of interest was death during hospitalization, defined as emergency department discharge disposition of "Death" or hospital discharge disposition of "Expired." The secondary outcome was the prevalence of complications at each center, when compared with the top 10 complications present in the aggregate data. Complication descriptions used were those defined in the NTDS.[10] To account for differences in baseline characteristics and injury severity of patients admitted to each trauma center, we used multivariate logistic regression modeling to allow for risk adjustment. Potential predictors of mortality were entered into the model, and a logit equation was derived based on the significant covariates. From the derived logit equation for predicted mortality, an expected mortality risk between 0 and 1 was calculated for each patient. Summing these expected mortalities for all patients at a given trauma center allowed an estimate of the expected total number of deaths for each hospital based on the covariate characteristics. The order of entry of a variable into the model is based on the c-index, representing the ability of that particular parameter to discriminate between survivors and nonsurvivors. For each trauma center the observed mortality rate (number of patients who died) was then divided by the calculated expected mortality rate to obtain an observed-to-expected (O/E) mortality ratio. These mortality ratios along with the 90% confidence intervals (CIs) allowed determination of trauma center performance in increasing order and identification of statistically significant high- or low-performance outliers. The 90% CI for each O/E ratio was calculated using Ulm's[11] method, which constructs the CIs based on the relationship between Poisson and the χ^2 distributions for observed events.

In some instances, specific incidents had missing values for potentially important covariates (Glasgow Coma Scale motor score, systolic blood pressure, and pulse rate). Because missing data are frequently not missing at random, we imputed these values using single imputation techniques. The final model and analysis included all the incidents that met TQIP entry criteria. To measure how well the model discriminates between prediction of death and survival, a c-index was calculated. The c-index typically takes on a value between 0.5 and 1.0, with a c-index of 0.5 representing a model that has no discrimination and is no better than a coin flip at determining outcome. A c-index of 1.0 represents perfect discrimination. The Hosmer and Lemeshow Goodness-of-Fit test was also computed to evaluate model calibration.

To evaluate the variability in rates of death across low, average, and high-outlier status trauma centers, we estimated the odds of death in each of the performance strata. The reference group consisted of those hospitals with low-outlier status, i.e., centers whose entire 90% CI was below 1 along with its O/E ratio. This analysis was performed for the aggregate, blunt multisystem injury, and single-system injury cohorts. It was not possible to conduct this analysis for the penetrating truncal injury cohort given the small number of patients and lack of any outlier status found within this group of patients.

Reports were compiled for each center and distributed to the pilot centers in January 2009. A sample benchmark report is available electronically at http://www.facs.org/trauma/ntdb/tqipsamplereport.pdf. Feedback on the first batch of benchmark reports was sought from trauma medical directors (or designate), registrars, and program managers at the participating trauma centers using an electronic survey instrument (SurveyMonkey.com, Portland, OR). This feedback was solicited in the form of yes/no questions and also allowed for detailed written comments in follow-up questions to negative responses.

RESULTS

Facilities participating in the TQIP pilot study reflect a range of Level I and II programs (Table 3). Although the majority of participants are Level I University-based trauma centers, efforts were made to include Level II centers to represent the experience of community trauma centers and to achieve geographic diversity. Application of the inclusion and exclusion criteria to the submitted NTDB data for the 23 participating centers allowed for creation of a TQIP data cohort referred to as the "aggregate" in subsequent tables. Of 43,701 trauma incidents for which data were submitted, a total of 15,801 patients met the inclusion criteria for TQIP analysis and reporting. The median number of incidents for a center was 550, and the range was from 192 to 1,479.

Each trauma center (identified as the "index hospital" in subsequent tables) received a report detailing their own patients and performance in comparison with all TQIP centers for both the aggregate population and for each of the three cohorts described above. Table 4 illustrates the TQIP patient characteristics for age, gender, ISS, injury type, and mechanism. Crude values for patient outcomes are reported in Table 5 along with their 90% CIs. Significant predictors of mortality that were kept in the final risk-adjustment model and their order of entry are listed in Table 6. The c-index for the final model was 0.901 and the Hosmer and Lemeshow Goodness-of Fit test was not significant ($p = 0.62$), indicating that the model fit the data well. The 8.7% of incidents were missing a Glasgow Coma Scale motor score, 3.6% were missing systolic blood pressure, and 3.4% were missing pulse rate. All these missing values were imputed to avoid biased exclusion of patients from the benchmark analysis.

TABLE 3. TQIP Facility Information

Variable	No. of Hospitals
Trauma center level	
I	17
II	6
Bed size	
≤200	0
200–400	5
401–600	6
>600	12
Teaching type	
Community teaching	7
Community nonteaching	1
University	15
Hospital type	
For profit	1
Not for profit	22
Region	
Northeast	5
Midwest	5
South	7
West	6

TABLE 4. TQIP Patient Characteristics

Characteristic	Index Hospital	Aggregate	Blunt Multisystem Injury	Penetrating Truncal Injury	Blunt Single-System Injury
Patients, N	442	15,801	2,874	1,238	11,689
Age (yr)					
16–25, N (%)	90 (20.4)	3,386 (21.4)	815 (28.4)	552 (44.6)	2,019 (17.3)
26–35	55 (12.4)	2,230 (14.1)	474 (16.5)	304 (24.6)	1,452 (12.4)
36–45	61 (13.8)	2,189 (13.9)	462 (16.1)	213 (17.2)	1,514 (13.0)
46–55	85 (19.2)	2,229 (14.1)	427 (14.9)	110 (8.9)	1,692 (14.5)
55–65	47 (10.6)	1,620 (10.3)	299 (10.4)	38 (3.1)	1,283 (11.0)
>65	104 (23.5)	4,147 (26.2)	397 (13.8)	21 (1.7)	3,729 (31.9)
Gender, N (%)					
Female	164 (37.1)	5,395 (34.1)	944 (32.8)	116 (9.4)	4,335 (37.1)
Male	278 (62.9)	10,402 (65.8)	1,930 (67.2)	1,122 (90.6)	7,350 (62.9)
Injury Severity Score, N (%)					
9–15	209 (47.3)	7,612 (48.2)	94 (3.3)	832 (67.2)	6,686 (57.2)
16–24	148 (33.5)	5,728 (36.3)	951 (33.1)	244 (19.7)	4,5333 (38.8)
>24	85 (19.2)	2,461 (15.6)	1,829 (63.6)	162 (13.1)	470 (4.0)
Injury type, N (%)					
Blunt	430 (97.3)	14,563 (92.2)	2,874 (100)	—	11,689 (100)
Penetrating	12 (2.7)	1,238 (7.8)	—	1,238 (100)	—
Mechanism, N (%)					
Motor vehicle traffic	174 (39.4)	6,817 (43.1)	2,138 (74.4)	0 (0.0)	4,679 (40.0)
Fall	139 (31.5)	5,499 (34.8)	430 (15.0)	0 (0.0)	5,069 (43.4)
Struck by or against	30 (6.8)	1,003 (6.4)	68 (2.4)	0 (0.0)	935 (8.0)
Firearm	7 (1.6)	665 (4.2)	0 (0.0)	665 (53.7)	0 (0.0)
Transport other	66 (14.9)	913 (5.8)	198 (6.9)	0 (0.0)	715 (6.1)
Cut/pierce	5 (1.1)	573 (3.6)	0 (0.0)	573 (46.3)	0 (0.0)

TABLE 5. TQIP Outcomes

Outcome	Index Hospital	Aggregate
All patients		
Crude mortality, % (90% confidence interval)	6.8 (4.9–9.1)	7.7 (7.4–8.1)
Blunt multisystem injury		
Crude mortality, % (90% confidence interval)	18.5 (12.9–25.3)	16.4 (15.3–17.6)
Penetrating truncal injury		
Crude mortality, % (90% confidence interval)	8.3 (0.4–33.9)	12.4 (10.9–14.1)
Blunt single-system injury		
Crude mortality, % (90% confidence interval)	2.3 (1.1–4.2)	5.1 (4.8–5.5)
Length of stay		
Median, days (IQR)	5 (3–9)	6 (3–10)
Mean, days (SD)	8.0 (9.95)	9.1 (12.2)
ICU length of stay		
Median, days (IQR)	3 (1–7)	3 (2–7)
Mean, days (SD)	5.5 (8.1)	6.4 (8.8)

The Clopper-Pearson confidence interval was used.
IQR, interquartile range; SD, standard deviation.

TABLE 6. Mortality Prediction Model

Variable	Odds Ratio	90% CI	c-Index
Initial GCS motor score in ED			
1	11.3	9.6–13.3	0.726
2–5	4.1	3.5–4.8	0.806
6	Reference		
Initial systolic blood pressure in ED			
0	16.3	9.1–29.4	0.811
1–90	3.6	3.0–4.4	0.827
>90	Reference		
Injury Severity Score			
>24	3.9	3.3–4.6	0.857
9–24	Reference		
Age (yr)			
>65	5.9	5.1–6.9	0.890
16–65	Reference		
Initial pulse rate in ED			
0–40 bpm	6.7	4.0–11.2	0.891
>40 bpm	Reference		
Mechanism of injury			
Firearm	4.3	3.2–5.7	0.895
All other mechanisms	Reference		
Head Injury Severity (AIS)*	1.4	1.3–1.5	0.899
Abdominal Injury Severity (AIS)*	1.3	1.2–1.4	0.900
Transfer status			
Transferred from an outside center	0.9	0.8–1.0	0.901
Transported from the field	Reference		

GCS, Glasgow Coma Scale; ED, Emergency Department.
*The increase in odds of death for each AIS of maximum severity categorized as follows: no injury, 1–2, 3–4, and 5–6.

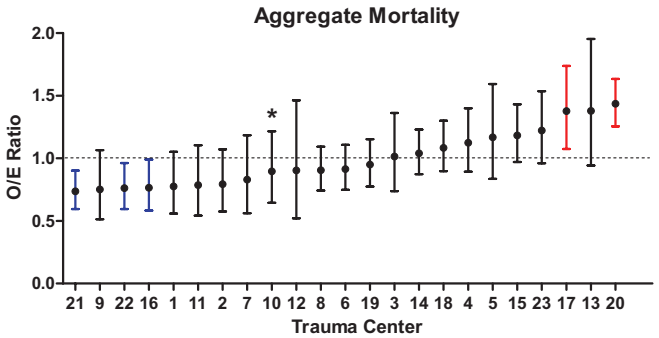

Figure 1. Risk-adjusted observed-to-expected mortality ratios for the aggregate group of all TQIP patients. The 90% confidence intervals shaded *blue* represent low-outlier status trauma centers, and those shaded *red* are high-outlier status centers. Confidence intervals that span one (shaded *black*) represent trauma centers whose risk-adjusted mortality is average. Trauma centers are identified by blinded numbers on the *x* axis. The index center is marked with an asterisk.

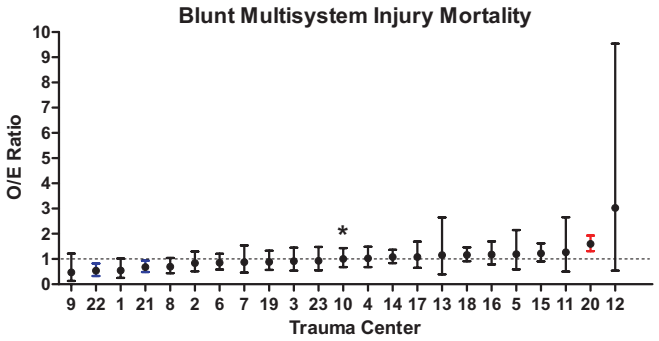

Figure 2. Risk-adjusted observed-to-expected mortality ratios for the blunt multisystem injury cohort. Note the change in rank for the index center and that trauma center 11 changes rank by 15 places.

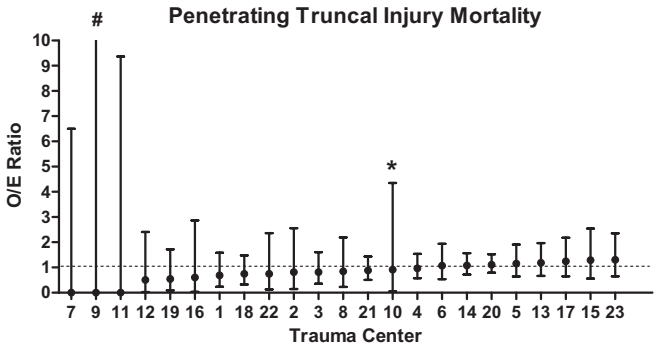

Figure 3. Risk-adjusted observed-to-expected mortality ratios for the penetrating truncal injury cohort. No centers were identified as high or low outliers. Three centers (7, 9, and 11) had no observed deaths.

The O/E mortality ratios are illustrated in Figure 1 for the total TQIP cohort. Trauma centers are listed in their order of rank from lowest to highest O/E ratio. Three centers were low outliers (blue highlighted 90% CI), and two centers were high outliers (red highlighted 90% CI). The remaining 18 centers demonstrated average mortality after risk adjustment.

Similar data are illustrated in Figures 2 to 4 for the three subgroups. For each individual cohort (blunt multisystem injury, penetrating truncal injury, and blunt single-system injury), the explanatory variables included in the model were the same as those listed in Table 6. However, the individual coefficients were recalculated for each subgroup patient cohort. The head AIS predictor was omitted from the model when estimating the number of deaths for the penetrating truncal injury cohort due to the low frequency of head injuries in this group. The rank order of trauma centers changes for the aggregate and each of the individual cohorts in Figures 1 to 4, indicating that there is not one center that excels at care delivery for all patients. The maximum change in rank was 15 places for one center between the aggregate and blunt multisystem injury cohort. The mean overall change in rank from the aggregate to any of the three subgroup cohorts was 3.6. Table 7 illustrates the impact of outlier status on the odds of death. When compared with a group of reference centers whom achieved low-outlier status both the average and high-outlier status hospitals had significantly elevated odds ratios for mortality, ranging from 1.6 to 5.9.

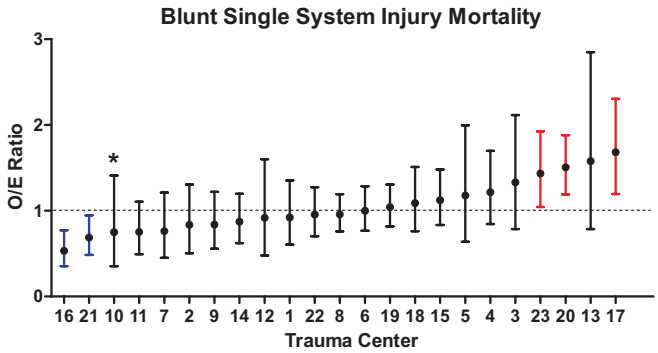

Figure 4. Risk-adjusted observed-to-expected mortality ratios for the blunt single-system injury cohort.

TABLE 7. TQIP Relative Risk of Mortality

Outlier Status	Aggregate	Blunt Multisystem Injury	Blunt Single-System Injury
Low	Reference	Reference	Reference
Average	1.6 (1.4–2.0)	2.1 (1.5–3.0)	2.6 (1.9–3.8)
High	3.3 (2.6–4.2)	5.0 (3.3–7.5)	5.9 (3.9–8.9)

A list of the top 10 complications and their prevalences is presented in Table 8. The response blank, not recorded, not done was also included within this list in addition to the actual recorded complications because it had the highest overall prevalence.

On receiving the reports, a total of 13 trauma registrars, 3 trauma program managers, and 8 trauma medical directors responded to our survey (Table 9). Comments were solicited and demonstrated how centers were using the data, e.g., "We have initiated a drill down of mortality rates among our patients with blunt single system injury because the data indicate that our mortality is higher than other centers for this group" and "Helps administration realize the value of the resources committed to trauma, let's us know we have no major problems with our trauma program." In addition, centers felt that the reports helped them to identify their own data completeness and quality issues.

DISCUSSION

These data demonstrate that a cohort of similar verified or designated trauma centers experience significant differences in their risk-adjusted mortality outcome for aggregate, blunt multisystem injury, and blunt single-system injury. The odds ratio of mortality varied substantially between the low-outlier and high-outlier trauma centers. In the aggregate group, this relative risk of mortality was 3.3, and for the single-system trauma cohort, the risk was 5.9 times higher for high-outlier facilities, when compared with the reference group of low-outlier hospitals. These TQIP findings from the first year of this pilot study confirm the previously published work of Shafi et al.[12,13] demonstrating variations in risk-adjusted mortality at similarly designated trauma centers using retrospective data from the NTDB and the Texas Emergency Medical Services/Trauma Registry. In

TABLE 8. TQIP Complications

NTDS Complication	Index Hospital	Aggregate	Blunt Multisystem Injury	Penetrating Truncal Injury	Blunt Single-System Injury
Patients, N	442	15,801	2,874	1,238	11,689
Blank/NR/ND, N (%)	325 (73.5)	9,167 (54.0)	1,469 (43.0)	758 (56.1)	6,940 (56.9)
No NTDS listed complications occurred, N (%)	0 (0.0)	5,115 (30.1)	808 (23.6)	346 (25.6)	3,961 (32.5)
Pneumonia, N (%)	42 (9.5)	614 (3.6)	281 (8.2)	44 (3.3)	289 (2.4)
Deep vein thrombosis, N (%)	9 (2.0)	249 (1.5)	111 (3.3)	16 (1.2)	122 (1.0)
ARDS, N (%)	4 (0.3)	231 (1.4)	96 (2.8)	19 (1.4)	116 (1.0)
Systemic sepsis, N (%)	10 (2.3)	226 (1.3)	117 (3.4)	26 (1.9)	83 (0.7)
Decubitus ulcer, N (%)	9 (2.0)	219 (1.3)	88 (2.6)	5 (0.4)	126 (1.0)
Cardiac arrest with CPR, N (%)	20 (4.5)	166 (1.0)	87 (2.5)	15 (1.1)	64 (0.5)
Coagulopathy, N (%)	46 (10.4)	162 (1.0)	87 (2.5)	31 (2.3)	44 (0.4)
Not applicable, N (%)	0 (0.0)	148 (0.9)	40 (1.2)	29 (2.2)	79 (0.7)
Drug or alcohol withdrawal syndrome, N (%)	4 (0.9)	146 (0.9)	31 (0.9)	6 (0.4)	109 (0.9)

NR, not recorded; ND, not done; ARDS, Adult (Acute) Respiratory Distress Syndrome.

TABLE 9. Survey

Questions	Yes (%)	No (%)
Was the report clear and understandable?	92	8
Is there additional information that you would like to see included in the report that is currently not included?	29	71
Was the information in the report useful for you and your facility?	90	10
Will you or your facility be taking any action regarding any of the information presented in the report?	63	37
Have you identified problems with the data submission related to your registry?	42	58

addition, substantial outcome rank changes were observed for trauma centers between the aggregate and each of the three cohorts indicating that no one trauma center excelled in all areas of trauma care delivery. Analysis for the penetrating truncal injury cohort was problematic because some centers had very few patients present in this cohort. Future analysis for penetrating truncal injury may need to be conducted over a longer time span or limited to trauma centers, which see a minimum volume of this type of injury.

TQIP performance reports will allow each individual center to compare their outcomes with other peer hospitals and to receive meaningful feedback on the quality of care provided to trauma patients. Some of the best and most respected quality improvement initiatives have involved local or regional collaborative efforts in which the participants are able to easily meet and communicate with each other. Examples are, the Northern New England Cardiovascular Disease Research Group,[14] MI Surgical Quality Collaborative,[15] and Blue Cross Blue Shield of Michigan Cardiovascular Consortium.[16] In the Northern New England Cardiovascular Disease Research Group, an intervention consisting of data feedback, training in continuous quality improvement techniques, and site visits resulted in a 24% decline in hospital mortality.[17] Within the state of Michigan, Blue Cross Blue Shield of Michigan Cardiovascular Consortium has evaluated the association of a continuous quality improvement initiative with practice and outcome variations of percutaneous coronary interventions.[16] After collection of baseline data, the continuous quality improvement intervention consisted of feedback on outcomes, working group meetings, site visits, selection of quality indicators, and use of bedside tools for quality improvement and risk assessment. Compared with baseline, the intervention group had higher use of preprocedural aspirin and glycoprotein IIb/IIIa receptor blockers, lower use of postprocedural heparin, and a lower amount of injected contrast media per case ($p < 0.05$). These changes were associated with lower rates of transfusions, vascular complications, contrast nephropathy, stroke, transient ischemic attack, and combined end points ($p < 0.05$). Darrell A. Campbell, Jr, one of the early disciples of NSQIP in the private sector, has recognized and promoted the concept that to maximize the benefits of NSQIP data, there has to be a mechanism—a local, face-to-face communication mechanism—where data can be evaluated, discussed, and turned into quality improvement.[15] Supporting this concept, MI Surgical Quality Collaborative has demonstrated a 13% reduction in the incidence of complications for inpatient general surgery cases during a 1-year interval. It is envisioned as TQIP is implemented and evolves, it could also support a similar framework of local or regional collaboration. This would likely fit within the structure of state and/or county trauma systems.

It is possible that factors other than quality of care may still influence the risk-adjusted mortality rates generated. The two current major limitations of TQIP and in interpreting the initial data fall into the areas of data quality and case ascertainment. Issues that have been identified and are being worked on include the following:

1. Data quality: These data were collected before the dissemination and implementation of NTDS data definitions and TQIP training of trauma registrars. It is possible that differences exist in data capture or coding of injury diagnosis. These differences might contribute to observed differences in the O/E mortality ratios or complication rates. Ongoing TQIP training programs, promulgation of NTDS data points and definitions, and internal/external data validation checks will serve to minimize the impact of these differences over time. Adherence to standardized definitions by all centers is critical to the benchmarking process.
2. Case ascertainment: We have no assurances that all patients who meet inclusion criteria for TQIP are included in the registry. It is plausible that selected high-risk (or low risk) patients might be systematically excluded from hospital registries. In the future, external validation with review of emergency department logs will assure that all who meet inclusion criteria are included in the analyses.
3. Selection bias: The current TQIP report was based on 23 trauma centers that volunteered to participate in the pilot study. These centers were not randomly chosen; hence, they may not represent the spread of all Level I or II trauma centers. Because more centers are recruited and brought on-line, TQIP will begin to more accurately reflect the broad scope of available trauma systems within the United States and Canada. Hence, the differences between those participating and those not participating will likely be reduced.
4. Performance over time: A trauma center's performance may vary over time. The current report presents a single snapshot in time. Consequently, the need for ongoing participation by trauma centers and continuous risk-adjustment model evolution to allow for constant PI assessment that is credible and reliable.
5. Chance: Statistical models are simply estimates. It is possible that chance alone led to the position of a center's performance relative to its peers. However, the likelihood of this occurrence by chance alone is less than 10% (based on a 90% CI). Again this emphasizes the necessity for ongoing participation by trauma centers and assessment of rank order changes from one report to the next.
6. In-hospital outcomes: O/E mortality ratios are based on in-hospital mortality. Differences in discharge disposition or access to alternate levels of care might influence in-hospital mortality rates.

The annual TQIP reports are intended for the exclusive use of the participating centers for self-evaluation and PI. It is important that these reports remain anonymous and are not used for marketing purposes or competitive advantage. This is a guideline that has been successfully adhered to by NSQIP and represents a strength of their nonpunitive approach to PI. It is not the intent of the COT to allow or consider TQIP to be a watchdog for the verification review committee. Rather, what is expected is that participating centers examine their own data by "popping the hood" when the "dashboard warning light goes off" (high-outlier status). It is even possible for a center to on occasion be a high outlier and not have quality of care issues. An example might be a center that during a select time period has a large amount of patients with potentially survivable injuries that undergo voluntary withdrawal of care secondary to patient or family wishes.

There is a necessity for additional work in the area of complications because some centers did not submit any NTDS defined complications. Identification and capture of the pertinent complications is an important area of future emphasis. Trauma patients experience markedly higher rates of almost every complication studied, when compared with NSQIP data.[18] Avoidance or a reduction in the severity of complications is important because complications have the potential to increase preventable mortality (failure to rescue effect), add to length of stay, and drive-up costs. Studies have quantified the costs savings for avoidance of major complications in the area of general surgery and trauma to be on the order of $12,000 to 40,000 per incident.[19,20] Strategies to improve the quality of complication data and feedback include better data definitions, elimination of dubious data fields (coagulopathy), emphasis on a select group of high-impact complications (pneumonia, sepsis, and deep venous thrombosis), and creation of composite measures. Implementation of external validation methods through site visits will also allow for ongoing education and identification of problems in this area.

The cost of providing a continuous quality improvement program is important when assessing its expense-to-benefit ratio. Participation in TQIP is currently priced at $9,000 per year, plus the cost of registrar salaries and trauma registry software. Because compilation and maintenance of a trauma registry is already mandated as a criterion for trauma center verification, the additional cost to a currently verified trauma center is really only the participation fee.[2] Therefore, TQIP has the potential to cost much less than NSQIP, which is priced at a fee of $35,000 per year plus the cost of at least one nurse per site to serve as the Surgical Clinical Reviewer ($70,000–$100,000).[21]

CONCLUSION

In summary, this pilot study has demonstrated the feasibility of TQIP as a collaborative quality improvement vehicle for trauma that uses existing resources and systems present at the local trauma center and national level. TQIP has the potential to cost significantly less than NSQIP and may be optimally positioned to yield quality improvement results through interventions because of the systems approach to trauma that is already underscored by the COT. To move TQIP to the next stage will require standardization, training, and validation. A philosophy change will also be necessary because the program cannot be all things to all people. Centers will have to exhibit flexibility and change in how they capture and record trauma registry data. The tradeoff for these changes is upside value in being able to create a local, regional, and national trauma system with risk-adjusted benchmarking that allows continuous assessment of performance by individual trauma centers and implementation of quality improvement interventions.

REFERENCES

1. American College of Surgeons (ACS Web site). Available at: http://www.facs.org/about/committees/index.html#Trauma. Accessed August 25, 2009.
2. American College of Surgeons. *Resources for Optimal Care of the Injured Patient*. Chicago, IL: American College of Surgeons, Committee on Trauma; 2006.
3. American College of Surgeons (ACS Web site). Available at: http://www.socialtext.net/data/workspaces/acs-demo-wiki/attachments/performance_improvement_and_patient_safety_reference_manual:20071008015138-1-16755/original/How-to%2520manual.pdf. Accessed August 30, 2009.
4. Champion HR, Copes WS, Sacco WJ, et al. The major trauma outcome study: establishing national norms for trauma care. *J Trauma.* 1990;30:1356–1365.
5. Khuri SF. The NSQIP: a new frontier in surgery. *Surgery.* 2005;138:837–843.
6. Khuri SF, Daley J, Henderson W, et al. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg.* 1997;185:315–327.
7. Khuri SF, Daley J, Henderson WG. The comparative assessment and improvement of quality of surgical care in the Department of Veterans Affairs. *Arch Surg.* 2002;137:20–27.
8. Khuri SF, Henderson WG, Daley J, et al. Successful implementation of the department of Veterans Affairs' National Surgical Quality Improvement program in the private sector: the Patient Safety in Surgery study. *Ann Surg.* 2008;248:329–336.
9. MacKenzie EJ, Sacco WJ. *ICDMAP-90 Software: User's Guide*. Baltimore, MD: Johns Hopkins University and Tri-Analytics; 1997:1–25.
10. National Trauma Data Standard of the NTDB (NTDB Web site). Available at: http://www.ntdsdictionary.org/dataElements/dataset Dictionary. Accessed December 4, 2009.
11. Ulm K. A simple method to calculate the confidence interval of a standardized mortality ratio (SMR). *Am J Epidemiol.* 1990;131:373–375.
12. Shafi S, Nathens A, Parks J, et al. Trauma quality improvement using risk adjusted outcomes. *J Trauma.* 2008;64:599–606.
13. Shafi S, Stewart RM, Nathens AB, et al. Significant variations in mortality occur at similarly designated trauma centers. *Arch Surg.* 2009;144:64–68.
14. O'Connor GT, Plume SK, Olmstead EM, et al. A regional prospective study of in-hospital mortality associated with coronary artery bypass grafting. The Northern New England Cardiovascular Disease Study Group. *JAMA.* 1991;266:803–909.
15. Campbell DA Jr. Quality improvement is local. *J Am Coll Surg.* 2009;209:141–143.
16. Moscucci M, Rogers EK, Montoye C, et al. Association of a continuous quality improvement initiative with practice and outcome variations of contemporary percutaneous coronary interventions. *Circulation.* 2006;113:814–822.
17. O'Connor GT, Plume SK, Olmstead EM, et al. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. The Northern New England Cardiovascular Disease Study Group. *JAMA.* 1996;275:841–846.
18. Hemmila MR, Jakubus JL, Wahl WL, et al. Detecting the blind spot: complications in the trauma registry and trauma quality improvement. *Surgery.* 2007;142:439–448.
19. Dimick JB, Chen SL, Taheri PA, et al. Hospital costs associated with surgical complications: a report from the private-sector National Surgical Quality Improvement program. *J Am Coll Surg.* 2004;199:531–537.

20. Hemmila MR, Jakubus JL, Maggio PM, et al. Real money: complications and hospital costs in trauma patients. *Surgery*. 2008;144:307–316.
21. American College of Surgeons, National Quality Improvement Program (ACS NSQIP Web site). Available at: http://acsnsqip.org/main/resources_faqs.asp. Accessed December 4, 2009.

DISCUSSION

Dr. L. D. Britt (Norfolk, Virginia): President Jurkovich, Secretary Cioffi, I want to applaud the Program Committee for putting together what I feel will be a superb program and thank them for allowing me to discuss this excellent and timely paper.

I also want to thank the authors for sending me the completed manuscript in a timely manner. Although this is one of a series of population-based investigations derived from robust, a robust National Trauma Databank, this is a pivotal and unique study designed to assess feasibility and acceptance of trauma quality improvement program prior to full implementation.

As the authors highlighted, ten institutions along with the American College of Surgeons participated in conducting this pilot study that included 23 ACS-verified or state-designated Level I and II trauma centers.

The authors should be commended for initiating such a project and being forthright in underscoring some of the salient limitations, including the obvious challenge in benchmarking morbidity and in benchmarking mortality following the penetrating injury due to the small sample size.

As highlighted by the authors, the mantra of the American College of Surgeons' Committee on Trauma has always been and always will be unwavering commitment to improving all phases of the management of the injured patient.

Demonstrating variations in risk-adjusted mortality and similarly designated trauma centers is important. However, I must ask the authors the difficult and most pressing questions.

Even though this was a pilot study, were there any substantive evidence, substantive evidence, that implementation of the TQIP has resulted in identifiable and documented quality improvement of the injured patient?

And if so, what would be or what would be the specific matrix outside of mortality that you would use to assess that?

The concept of quality care has three components: structure, process and outcomes. While we continue to appropriately address the first two, I submit that not enough emphasis is being placed on definitively determining if intervention is actually affecting outcome.

Again, I commend the authors for what will ultimately be considered a landmark, pivotal publication. Thank you.

Dr. Mark R. Hemmila (Ann Arbor, Michigan): We have data from the first year and we have fed the benchmark reports back.

It's hard to have interventions at this point. We fully intend to have interventions and we intend to follow the same principles that other successful CQI initiatives have followed.

I can only point to the example of the New England Cardiovascular Disease group in which it took considerable time for them to actually collect the data, feed it back, and implement intervention. They were then able to see a constant down trending of their mortality rates over a period of many years.

What we have is, unfortunately, a very slow process but a very robust process in which we're trying to bring everybody onto the same page with standardization.

Once we have that in place we will have a very powerful tool in which to answer important questions. So I think it's early to pass judgment on the success or not of TQIP.

I think you need to give it a chance to grow, give it a chance to mature and see what the true results are over time.

Dr. Samir Fakhry (Charleston, South Carolina): I agree with Dr. Britt. You're setting the stage for a vital component of our work in the future and we look forward to joining you next year. Hopefully you've received the check from us.

The question that continues to bother me in all the work with databases over the years is how do we ensure appropriate capture of not so much mortality, which is a rather discreet endpoint, but the complications. I think you're doing wonderful work defining the complications and adopting a standardized dataset. How do we make sure that our centers capture complications in a systematic manner that truly reflects what is going on?

Will you adopt a similar structure as what NSQIP has done or what other programs around the country have done? Thank you very much. And, once again, I commend you on this excellent work.

Dr. George Velmahos (Boston, Massachusetts): I think at the heart of the matter is risk adjustment. Whether the dataset is intended to compare and profile centers, it's almost irrelevant, it will happen anyway.

So, the question is how do we accurately adjust for risk? And beyond some simple criteria like age and injury severity score, there is so many points that are needed for accurate risk adjustment: transfer status, in-house call, so many things.

Could you tell us a little bit more about risk adjustment in this study?

Dr. Mark R. Hemmila (Ann Arbor, Michigan): To answer Dr. Fakhry's question, I think we're very concerned about the capture of the deaths and complications. That certainly is the numerator.

And we're worried about the denominator in terms of the total number of patients. It's important for us to be transparent with our inclusion and exclusion criteria.

It's also important for us to carry out external validation through site visits which will be occurring. This will be an opportunity for us to verify the data collection as well as provide education to the centers and their registrars on the standardization process.

I think the only assurance that we can give you is that this has worked out well for NSQIP. It will be a learning experience for everyone. We are well aware that some centers feel that they could be penalized, so-to-speak, for having a close proximity to the scene of the injury whereas other centers have long transport times. Therefore, some people who would die in route or never make it to them don't show up in the data.

Another important thing that sort of answers Dr. Velmahos' question as well is that this is your data. This is for your use. It's not intended to be a score card that goes out to the world.

You all put in a tremendous amount of resources for your trauma registry in terms of time and money and don't necessarily get back a lot for that effort. I'd like to say that you could invest that money and get much better feedback by drawing everybody onto the same page and trying to provide risk adjusted feedback on outcomes.

In regard to the risk adjustment question, it turns out there are actually very few variables that are really relevant. Dr. Nathens presented a paper I believe a year or two ago where he categorized the change that occurred in rank based on which risk-adjustment variables were included.

Once you get past the first four to five co-variates there was very minimal change in rank, less than 1 percent. It turns out that we can do a pretty credible job with a fairly small model which works in our favor since it would be very cumbersome to collect a lot of co-morbidity or lab values for trauma.

Dr. Gerard J. Fulda (Newark, Delaware): Just to follow up with Dr. Britt's comments about outcome, one of the experience with the NSQIP program that we've had is that with the data being blinded with respect to your performance, it becomes difficult to take, when you become an outlier, to identify the senders that are on the other end of the spectrum, outliers either good or bad, and getting together to identify what makes you a high outlier or a lower outlier.

What plans do you have or processes do you want to put into place to either un-blind the data among certain centers or to get cooperation between the high and low outliers? Thank you.

Dr. Palmer Q. Bessey (New York, New York): I was interested that of all the variables you have here, there is nothing that really relates to the centers as being any different other than transfer, I believe. Otherwise, they're all basically patient level variables.

And I wonder if you've accounted for variables, variability within the centers either with hierarchal modeling, have you looked at things such as size that we think make a big difference in terms of experience and those aspects of it? Otherwise, it's very nice work. Congratulations.

Dr. Mark R. Hemmila (Ann Arbor, Michigan): So to answer the first question about the outcomes, what has worked for many groups is to have a meeting where you get together and basically share the information.

Certainly if you don't wish to identify yourself as a high or low outlier, that's your business. But it turns out you have a chance to work with your friends, potentially, to improve your performance and share your successes.

This has been borne out pretty well in Michigan through an initiative that Skip Campbell has been spearheading where he looks at NSQIP in the State of Michigan (Michigan Surgical Quality Collaborative).

There is a graph where it shows morbidity for NSQIP as a whole, a fairly straight line, not much change, but in Michigan because they have a regional collaborative in which these hospitals get together, there is a nice down trending of the O to E ratios.

I think that's the next important key in this is once you provide the information back, that you have an exchange of information so that people can learn from each other.

In terms of the center characteristics, it is possible to enter those in, as you said, with a hierarchal model.

One of the problems is that, to use a baseball analogy, you can go through and do all that, create a model, but it doesn't account for things that happen in the future so you don't have the fact that a patient got injured.

In baseball you can predict what somebody's batting average will be based on previous things but it doesn't take into account things like injury or the fact that the patient's wife cheated on him or something that affects his mental status.

So there are only so many things you can do. I think Wayne Meredith said it best when he said, "We're not trying to have the perfect model, we're just trying to be credible and reliable."