Using Machine Learning to Identify Change in Surgical Decision Making in Current Use of Damage Control Laparotomy

John A Harvin, MD, MS, FACS, Charles E Green, PhD, Claudia Pedroza, PhD, Jon E Tyson, MD, MPH, Laura J Moore, MD, FACS, Charles E Wade, PhD, John B Holcomb, MD, FACS, Lillian S Kao, MD, MS, FACS

BACKGROUND: In an earlier study, we reported the successful reduction in the use of damage control laparotomy (DCL); however, no change in the relative frequencies of specific indications was observed. In this study, we aimed to use machine learning to help identify the changes in surgical decision making that occurred.

STUDY DESIGN: Adult patients undergoing emergent trauma laparotomy were included: pre-quality improvement (QI): January 1, 2011 to October 31, 2013 and post-QI: November 1, 2013 to June 30, 2016. Using 72 variables before or during emergent laparotomy, random forest algorithms predicting DCL before and after a QI intervention were created. The main end point of the algorithms was the strength of individual factor significance in predicting the use of DCL, calculated by determining the mean decrease in accuracy (MDA) in the model if that variable was removed.

RESULTS: In the pre-QI group, 24 of 72 factors significantly predicted DCL, the strongest being bowel resection (mean MDA 16) and operating room RBC transfusions (mean MDA 15). The remaining variables were spread along the continuum of care from injury to emergent laparotomy end. In the post-QI group, 12 of 72 factors significantly predicted DCL, the strongest being last operating room lactate (mean MDA 12) and operating room RBC transfusions (mean MDA 14). In addition to having 12 fewer significant factors predictive of DCL, the predictive factors in the post-QI group were mainly intraoperative factors.

CONCLUSIONS: A machine learning analysis provided novel insights into the changes in decision making achieved by a successful QI intervention and should be considered an adjunct to understanding successful pre- and post-intervention QI studies. The analysis suggested a shift toward using mostly intraoperative factors to determine the use of DCL. (J Am Coll Surg 2019;228:255−264. © 2019 by the American College of Surgeons. Published by Elsevier Inc. All rights reserved.)

In a learning trauma system, the use of an innovation is continuously and iteratively refined based on evolving care needs and circumstances.[1] Rapid learning occurs with real-time access to data, efficient data analytics, and a culture of learning. Machine learning, which is increasingly being used in healthcare, is a data analytic strategy that is ideally suited to a learning trauma system. Machine learning is a software-based strategy by which factors involved in surgical decision making can be quantified and future decisions can be predicted.[2] Supervised machine learning uses historical data to train models to predict outcomes in new cases.

Disclosures Information: Nothing to disclose. Disclosures outside the scope of this work: Dr Wade is a paid consultant to Haemonetics, receives grants money from Masimo Grifols, and receives royalties and stock options from Decisio, LLC. Dr Holcomb is Chief Medical Officer for Paytime Medical.

Support: This work was supported by the Center for Clinical and Translational Sciences, which is funded by National Institutes of Health Clinical and Translational Award UL1 TR000371 and KL2 TR000370 from the National Center for Advancing Translational Sciences.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Advancing Translational Sciences or the National Institutes of Health.

Received October 31, 2018; Revised December 26, 2018; Accepted December 27, 2018.

From the Department of Surgery (Harvin, Moore, Wade, Holcomb, Kao), Center for Translational Injury Research (Harvin, Moore, Wade, Holcomb), Department of Pediatrics (Green, Pedroza, Tyson), and Center for Clinical Research and Evidence Based Medicine (Green, Pedroza, Tyson, Kao), University of Texas McGovern Medical School, Houston, TX.

Correspondence address: John A Harvin, MD, MS, FACS, Department of Surgery, University of Texas McGovern Medical School, 6431 Fannin St, MSB 4.264, Houston, TX 77030. email: john.harvin@uth.tmc.edu

https://doi.org/10.1016/j.jamcollsurg.2018.12.025
ISSN 1072-7515/19

Abbreviations and Acronyms

DCL = damage control laparotomy
ED = emergency department
MDA = mean decrease in accuracy
OR = operating room
QI = quality improvement

Damage control laparotomy (DCL) is an innovation from the 1990s in which an abbreviated laparotomy is performed to achieve source control of hemorrhage and contamination after trauma. The patient is left with an open abdomen, resuscitated to normal physiology in the ICU, and returned to the operating room (OR) for definitive reconstruction once clinically stable. Initially described in a small subgroup of coagulopathic patients requiring packing and massively transfused patients with a major abdominal vascular injury, DCL improved survival significantly.[3,4] However, DCL comes at the price of considerable morbidity, such as enterocutaneous fistulas, complex ventral hernias, and infectious complications.[5,6] Damage control laparotomy also commits patients to at least 1 reoperation and is associated with increased resource use and costs.[7] Despite these downsides, the indications for DCL have increased exponentially over time, without high-quality evidence demonstrating benefit.[8,9]

In an earlier study, we reported the success of a single-center, multifaceted quality improvement (QI) intervention in reducing the rate of DCL without increasing morbidity or mortality.[10] The intervention included consensus building about indications for definitive laparotomy and DCL, as well as audit and feedback.[11] The QI intervention resulted in an immediate and sustained reduction in DCL from 38% to 23%. Although the number of agreed on indications for DCL decreased, there was no change in the relative frequencies of indications for DCLs performed before and after the QI intervention. Explicit knowledge of the critical factors and conditions that resulted in the decreased use of DCL was lacking; this is a common shortcoming of QI research.[12]

The purpose of this study was to use supervised machine learning to provide insight about how a QI intervention changed surgical decision making. We hypothesized that machine learning could be used to identify a change in the factors predicting DCL as a result of the QI intervention. An explicit understanding of how surgical decision making changed is necessary to inform future evidence-based care.

METHODS

The study was an exploratory analysis of a pre-/post-intervention QI project performed at the Red Duke Trauma Institute at Memorial Hermann Hospital-Texas Medical Center, an American College of Surgeons-verified Level I trauma center that is the primary teaching hospital for the McGovern Medical School at UTHealth. The Red Duke Trauma Institute is 1 of only 2 adult Level I trauma centers in Houston, TX. The IRB deemed the project to be QI and exempt from review.

Participants of this study were surgeons performing emergent laparotomies on adult trauma patients older than 15 years of age. Emergent laparotomy was defined as direct admission to the OR from the emergency department (ED). Patients who died intraoperatively were excluded.

The cohort was divided into patients undergoing emergent trauma laparotomy during 2 time periods: pre-QI from January 1, 2011 to October 31, 2013 and post-QI from November 1, 2013 to June 30, 2016. Patient data for the pre-QI group were retrospectively collected for 2 peer-reviewed studies and was the same cohort as the control group of the QI study published previously[5,10]; the data for the post-QI group were collected prospectively using the same definitions and methods and included the QI cohort of the QI study published previously, plus an additional 8 months of accrued patients.

A total of 72 variables were included in the analysis and were obtained throughout the continuum of care—from admission (patient demographic and injury information) through the ED (vital signs, laboratory results, and resuscitation volumes) and through the OR (first and last vital signs and laboratory results, OR resuscitation, estimated blood loss, procedures performed, and surgeon). Surgeons with fewer than 10 laparotomies in the individual time periods were grouped together for the analysis. The end point predicted was performance of a DCL.

Random forest algorithm: a machine learning tool

Random forest is a computer learning algorithm based on de-correlated decision trees.[13] This methodology was chosen over other quantitative methods because it is robust to correlated predictors and potentially nonlinear relationships between the predictors and the criterion variables. First, a portion of the data is randomly selected to be used as a training set. The random forest algorithm is developed on the training set and then applied to the remainder of the data, which are referred to as the test set.

To develop the forest, the individual trees must be grown. To grow the tree, each variable at each decision node along the decision tree is selected randomly from a prespecified number of candidate variables. From these candidate variables, the algorithm randomly identifies the best variable and cut point on that variable. This continues iteratively over and over using the training set to minimize misclassification in the resulting branches until the tree is fully developed. The entire process is then repeated for the number of trees specified in the forest.

The most important parameter governing tree performance is the number of candidate variables selected at decision node.[14] This number is empirically determined using 5-fold cross-validation in the training set.[15] Based on the misclassification performance of multiple random forests on the training set, the number of candidate variables available at each decision node is selected to use in the random forest for the test set.

The optimized random forest of *n* trees is then applied to the test set. Individuals in the test set are given predicted categories based on the results of the *n* trees. The distribution of predicted outcomes values is then compared with known outcomes values and the model accuracy and sensitivity can be calculated.

Variable significance is calculated by determining the mean decrease in accuracy (MDA) in the model if that

Table 1. Group Characteristics in the Pre-Quality Improvement and Post-Quality Improvement Periods

Characteristics	Pre-QI group* (n = 559†)	Post-QI group‡ (n = 582†)	p Value
All emergent laparotomy, n	582	611	<0.001
Definitive laparotomy, n (%)	339 (58)	455 (74)	—
DCL, n (%)	220 (38)	127 (21)	—
Intraoperative death, n (%)	23 (4)	29 (5)	—
Relative frequency of indication			
DCL, n	220	127	—
Packing, n (%)	134 (61)	76 (60)	0.28
Second look, n (%)	29 (13)	20 (16)	—
Hemodynamic instability, n (%)	30 (14)	17 (13)	—
ACS prophylaxis, n (%)	3 (1)	6 (5)	—
Contamination, n (%)	7 (3)	1 (1)	—
Expedite CT/ICU, n (%)	14 (6)	5 (4)	—
Expedite IR, n (%)	1 (0)	2 (2)	—
Unclear, n (%)	2 (1)	0 (0)	—
Demographics			—
Age, y, median (IQR)	34 (24 to 48)	32 (23 to 46)	0.16
Sex, male, n (%)	423 (76)	446 (77)	0.70
Penetrating mechanism, n (%)	210 (38)	288 (50)	<0.001
BMI, kg/m^2, median (IQR)	26 (24 to 30)	27 (24 to 31)	0.14
Emergency department vital signs and labs			
Temperature, °F, median (IQR)	97.6 (96.7 to 98.3)	97.7 (97.0 to 98.3)	0.047
Systolic blood pressure, mmHg, median (IQR)	111 (90 to 130)	115 (93 to 130)	0.09
Heart rate, beats/min, median (IQR)	99 (82 to 15)	100 (83 to 117)	0.59
Glasgow Coma Scale, median (IQR)	15 (6 to 15)	15 (12 to 15)	0.007
Base excess, mmol/L, median (IQR)	−3 (−7 to −1)	−4 (−7 to −1)	0.21
Injury severity, median (IQR)			
Head AIS	0 (0 to 2)	0 (0 to 0)	0.063
Chest AIS	2 (0 to 3)	2 (0 to 3)	0.51
Abdominal AIS	3 (2 to 4)	3 (2 to 4)	0.043
Injury Severity Score	19 (11 to 34)	18 (10 to 29)	0.035
Outcome, n (%)			
Mortality	52 (9%)	65 (11%)	0.30

*January 1, 2011 to October 31, 2013.
†The number for each column is total emergent laparotomies minus intraoperative deaths, which were excluded.
‡November 1, 2013 to June 30, 2016.
AIS, Abbreviated Injury Scale; ACS, abdominal compartment syndrome; DCL, damage control laparotomy; IR, interventional radiology; QI, quality improvement; IQR, interquartile (25^{th} to 75^{th}) range.

Table 2. Complete Case and Imputed Model Characteristics of the Pre-Quality Improvement and Post-Quality Improvement Groups

	Pre-QI group*		Post-QI group†	
Characteristic	Complete case model (n = 236)	Imputed model (n = 559)	Complete case model (n = 209)	Imputed model (n = 582)
Accuracy	0.772	0.857	0.822	0.873
Sensitivity	0.800	0.850	0.859	0.873
Specificity	0.684	0.873	0.600	0.875
Positive predictive value	0.889	0.930	0.929	0.981
Negative predictive value	0.520	0.743	0.409	0.488
AUC of ROC	0.812	0.921	0.804	0.896

*January 1, 2011 to October 13, 2013.
†November 1, 2013 to June 30, 2016.
AUC, area under the curve; QI, quality improvement; ROC, receiver operating characteristic.

variable was removed. As the MDA of a variable increases, the relative importance of that variable to the model increases.

Model creation

For both the pre-QI and post-QI groups, 2 models predicting DCL were created using random forest methodology (R Studio, version 1.0.143) for a total of 4 models. The 2 models in each time period included a model using complete data only (patients with missing data excluded) and a model using imputation to account for missing data. The 2 models acted as a sensitivity analysis to address variable missingness.[16]

Imputation was performed using random forest methodology in R. First, all missing continuous variables were replaced with the variable median value and all factor variables were replaced with the most variable frequent value. Next, a random forest was run on the completed data set and a proximity matrix was created and used to update the imputed values. For continuous variables, the imputed value was then updated using the scaled proximity score of the non-missing observations as a weighing factor. For factor variables, the imputed value was the category with the largest average proximity. This process was iterated 5 times.

The 4 models were created using a similar methodology. First, a training set was created using a random sample of two-thirds of the data. The most important parameter to tune in the random forest algorithm is the number of randomly selected variables evaluated for classification at each branch in the tree. Tuning this parameter used k = 5-fold cross-validation. The model that optimized the misclassification rate using k-fold cross-validation on the training set was then evaluated on a test set (random one-third hold-out sample).

A confusion matrix and receiver operator curve were then created to describe the accuracy of the model performance on the test data. Finally, the MDA, which is a measure of how many SDs that a raw score is above or below the population mean, for each variable was calculated. As the MDA of a variable increases, the relative importance of that variable to the model increases. In this analysis, the MDA was scaled to create a z-score to allow for comparison across variables.

Determination of significant variables

To fulfill the aim of the project—to describe how surgeon judgment changed after institution of the successful QI project—2 separate analyses were performed.

Variable significance in both periods

First, the MDA of variables in both models in both time periods were analyzed. Variables were considered to be significant if both models concurred that the variable MDA was either ≥ 2 or ≤ -2 (in the same direction). A cutoff of 2 SDs from the mean without correction for multiplicity was chosen, as we did not want to prematurely exclude any potentially modifiable factors in surgical decision making that could be used to improve and generalize this QI effort.

Variables with an MDA meeting the threshold of 2 or −2 were then listed in a side-by-side table comparing the pre-QI and post-QI periods. This method of analysis allowed the comparison of significant variables in both time periods as a way to understand what variables changed in the determination to perform DCL.

Net mean decrease in accuracy

Second, the difference in the MDA in complete cases in the pre-QI and post-QI groups was calculated for each variable. This method of analysis was performed to attempt to express both the vector (increase or decrease in importance) and the magnitude of the change in variable MDAs during the 2 time periods. A net difference in MDA was similarly derived from the 2 imputed models for all variables regardless of individual MDA. If the net MDA in both models was ≥2 or ≤ −2, the variable was considered to have had a significant change between the pre-QI and post-QI periods.

RESULTS

In the pre-QI group, 12,922 patients were admitted to the trauma center; 582 (5%) patients underwent emergent laparotomy, with a DCL rate of 38% (Table 1). In the post-QI group, 13,061 patients were admitted to the trauma center; 611 (5%) patients underwent emergent laparotomy, with a DCL rate of 21%. The groups had similar baseline demographics and injury severity, with the exception of a higher percentage of patients experiencing penetrating trauma in the post-QI group (Table 1). The relative frequencies of the stated indications for DCL were not statistically significantly different between the 2 time periods ($p = 0.28$). Intraoperative deaths occurred in and were excluded from the analysis in 23 (4%) and 29 (5%) of the pre-QI and post-QI groups, respectively.

Model accuracy

For both the pre-QI and post-QI groups, the accuracy of all models was good (range 0.772 to 0.857) (Table 2).

Table 3. Mean Decrease in Accuracy of Significant Variables Predicting Damage Control Laparotomy in the Pre-Quality Improvement and Post-Quality Improvement Periods

	Pre-QI group*			Post-QI group†		
Variable	Complete case model	Imputed model	Average	Complete case model	Imputed model	Average
MDA of demographic and injury variable						
Abdominal AIS	2	6	4	—	—	—
Injury Severity Score	2	8	5	4	6	5
Large bowel injury	3	4	3.5	—	—	—
Liver injury	5	4	4.5	—	—	—
Kidney injury	—	—	—	4	2	3
Pelvic fracture	4	4	4			
MDA of emergency department variable						
Temperature	6	9	7.5	—	—	—
Systolic blood pressure	6	9	7.5	4	6	5
Base excess	3	9	6	—	—	—
RBC transfusion	5	7	6	—	—	—
MDA of operating room variables						
First pH	6	7	6.5	—	—	—
First base excess	4	4	4	—	—	—
First lactic acid	7	8	7.5	7	8	7.5
RBC transfusion	9	21	15	14	15	14.5
FFP transfusion	7	15	11	15	15	15
Platelet transfusion	4	7	5.5	6	8	7
Estimated blood loss	4	13	8.5	6	9	7.5
Last systolic blood pressure	2	5	3.5	4	6	5
Last temperature	5	3	4	7	5	6
Last pH	6	3	4.5	—	—	—
Last base excess	6	5	5.5	—	—	—
Last lactic acid	6	13	9.5	11	14	12.5
MDA of procedure performed						
Any enteric resection	2	31	16.5	2	3	2.5
Large bowel resection	6	9	7.5	—	—	—
Hepatorrhaphy	7	5	6	—	—	—

Values represent the MDA of each variable scaled to a z-score. Variables were considered significant if the MDA was >2. A larger MDA represents a larger decrease in the accuracy of the model if that variable was removed, meaning the variable is highly predictive of the end point (ie damage control laparotomy).
*January 1, 2011 to October 13, 2013.
†November 1, 2013 to June 30, 2016.
AIS, Abbreviated Injury Scale; FFP, fresh-frozen plasma; MDA, mean decrease in accuracy; QI, quality improvement.

The complete case models had <50% of the entire cohort of each time period reflecting significant missingness of clinical variables (pre-QI complete case model contained 236 of 559 patients; post-QI complete case model contained 209 of 582 patients). The variables with the most missingness in the pre-QI group were ED base excess (17%), last OR pH (21%), last OR base excess (21%), and last OR lactic acid (21%). The variables with the most missingness in the post-QI group were last OR pH (21%), last OR base excess (21%), and last OR lactic acid (21%).

Variable significance in both periods

In the pre-QI group, there were 24 variables predicting DCL. Although the majority were intraoperative variables, demographics, injury characteristics, ED variables, and procedures performed also influenced the decision for DCL (Table 3). The variables with the largest MDA in both pre-QI models indicating a significant association with DCL were the performance of a bowel resection, ED systolic blood pressure, and OR transfusions (RBC and fresh-frozen plasma transfusions).

In the post-QI group, there were half as many variables predicting DCL. All but 3 of the variables were intraoperative variables. The most influential variables within the post-QI models were last OR lactic acid and OR transfusions (RBC and fresh-frozen plasma transfusions).

Although there were some—albeit noticeably fewer than in the pre-QI group—variables related to injury, ED factors, and OR procedures, the strongest predictors of DCL in the post-QI group were OR transfusions and physiology at the end of the case. In the pre-QI group, variables predictive of DCL were spread across the entire continuum of care from admission to the end of the OR. In the post-QI group, variables most predictive of DCL were OR factors (Fig. 1).

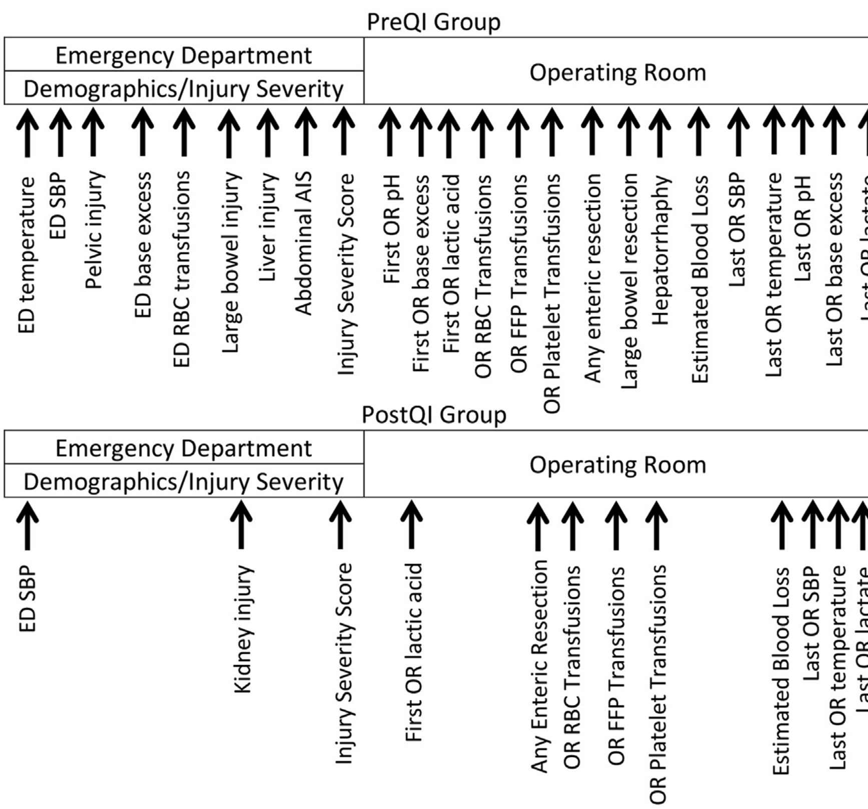

Figure 1. The timing of variables found to be significant predictors of damage control laparotomy before and after the quality improvement intervention. Prior, predictive variables were spread across the entire continuum of care, from injury to the end of the laparotomy. After, predictive variables are more concentrated on intraoperative factors, specifically transfusion volumes and end of operation physiology. AIS, Abbreviated Injury Scale; ED, emergency department; FFP, fresh-frozen plasma; OR, operating room; SBP, systolic blood pressure.

Table 4. Net Difference in Mean Decrease in Accuracy Between the Pre-Quality Improvement and Post-Quality Improvement Groups

Variable	Net complete case model	Net imputed model	Average net difference	Interpretation
Demographic and injury variable				
Penetrating injury	3	3	3	Decrease in importance
BMI	4	3	3.5	Decrease in importance
External Abbreviated Injury Scale	2	5	3.5	Decrease in importance
Liver injury	6	4	5	Decrease in importance
Kidney injury	−6	−2	−4	Increase in importance
Large bowel injury	2	3	2.5	Decrease in importance
Mesentery injury	−4	−2	−3	Increase in importance
Pelvis injury	4	5	4.5	Decrease in importance
Vascular injury	−2	−2	−2	Increase in importance
Emergency department variable				
Temperature	7	2	4.5	Decrease in importance
Systolic blood pressure	7	2	4.5	Decrease in importance
RBC transfusions	5	5	5	Decrease in importance
Operating room variable				
Surgeon	−9	−2	−5.5	Increase in importance
Last temperature	−2	−2	−2	Increase in importance
Last base excess	6	5	5.5	Decrease in importance
Procedure performed				
Hepatorrhaphy	6	3	4.5	Decrease in importance
Large bowel resection	4	9	6.5	Decrease in importance
Nephrectomy	−6	−2	−4	Increase in importance

The net difference in mean decrease in accuracy between the pre-quality improvement and post-quality improvement groups provided an estimate of the degree of change in surgical decision making and the vector of that change. For example, large bowel resection had the largest decrease in the importance of predicting the use of damage control laparotomy and surgeon had the largest increase in the importance of predicting the use of damage control laparotomy.

Net mean decrease in accuracy

Based on net MDAs (Table 4), the variables with the greatest decrease in importance from the pre-QI to the post-QI period were pelvic fracture, liver injury, ED temperature, ED systolic blood pressure, ED RBC transfusions, large bowel resection, and last OR base excess (Fig. 2). The variables with the greatest increase in importance were surgeon, kidney injury, and nephrectomy.

DISCUSSION

Supervised machine learning identified and quantified factors leading to the changes in surgical decision making for DCL before and after implementation of a successful QI intervention. Although the factors that contributed most to the effectiveness of the intervention were unknown previously, machine learning identified shifts in the timing of the decision for DCL and in the emphasis placed on specific physiologic variables, as well as associated injuries. Although the frequency of use of the broad indications for DCL did not change, the parameters for defining those indications became more specific.

The decision of whether or not to perform DCL can be challenging for multiple reasons. First, although there are observational studies supporting its use,[3,4] there are no completed randomized trials demonstrating benefit. In fact, a matched analysis of patients undergoing DCL vs definitive laparotomy at our trauma center suggested that DCL is associated with increased major abdominal complications, although the study was underpowered to identify a statistically significant difference.[17] A study at another major trauma center reported that delay of additional interventions, such as vascular procedures associated with DCL, resulted in increased mortality, unplanned re-explorations, and re-exploration for hemorrhage control.[18] Second, there is significant uncertainty in the literature and among surgeons about the indications for DCL. A scoping review of the literature identified 1,099 indications for damage control operation.[9] A subsequent content analysis by international experts in trauma surgery cited 101 appropriate indications for DCL.[8] Lastly, DCL is associated with increased resource use and costs. Higa and colleagues[7] demonstrated that a 78% reduction in DCLs resulted in

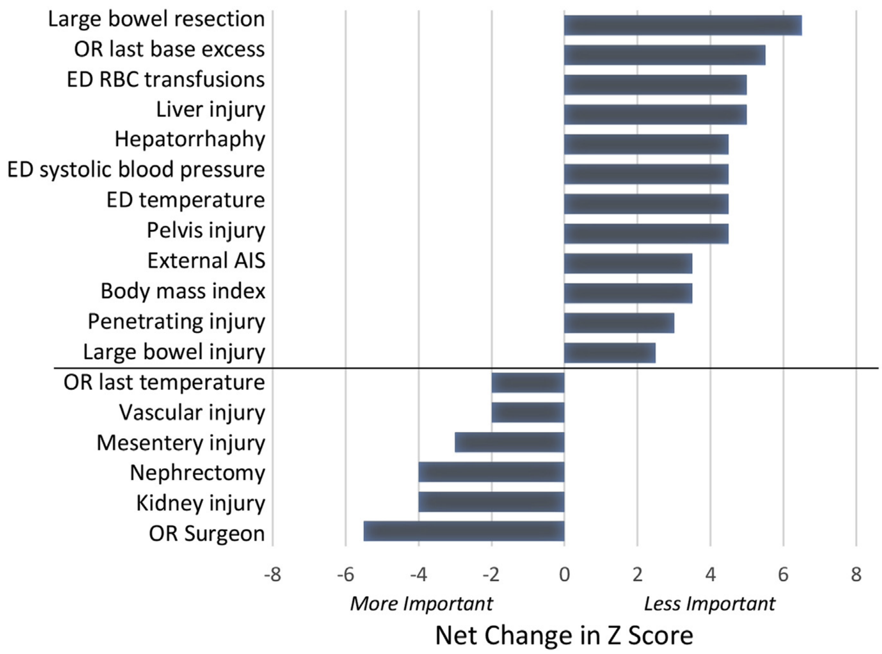

Figure 2. The net change in variables before and after the quality improvement intervention. There was a general de-emphasis on emergency department variables and bowel resections. The greatest increase in importance was the individual surgeon, likely indicating that the surgeon-specific rates of damage control laparotomy became more disparate after the quality improvement project. AIS, Abbreviated Injury Scale; ED, emergency department; OR, operating room.

a projected savings of $2.2 million in costs, without an increase in mortality.

Although this study was conducted at a single center, our pre-intervention decision making closely mirrored that of the trauma community. In an international cross-sectional survey of 366 surgeons, 15 preoperative and 23 intraoperative indications were deemed appropriate to justify damage control operation.[19] Before the QI intervention at our institution, 9 preoperative and 16 intraoperative factors were associated with the decision to perform DCL, including demographic, injury, ED, OR, and procedural variables. Several local influencing factors were similar to those cited in the international survey, such as preoperative temperature, systolic blood pressure, intraoperative pH, and estimated blood loss. After the QI intervention, our consensus-based indications were very similar to the highest-rated indications identified from the literature: injury pattern identified during operation, inability to control bleeding by conventional methods, amount of resuscitation provided, degree of physiologic insult, and need for staged abdominal or thoracic wall reconstruction (ie avoidance or treatment of abdominal compartment syndrome).[8,10]

The QI intervention resulted in a de-emphasis on preoperative factors and an increased reliance on the patient's status at the end of the operation for determination of the degree of physiologic insult and the adequacy of resuscitation. In particular, there was a significant decrease in significance attributed to specific physiologic parameters, such as last OR base excess. This shift correlates with a change in indication for DCL in our group from persistent acidosis alone to that associated with ongoing transfusion and continuous vasopressor infusion for hemodynamic instability. There was also a change in use of DCL for specific injury patterns and attempted bleeding control; liver injury and hepatorrhaphy were significantly associated with DCL before but not after the QI intervention. Although there was no change in the rate of packing (pre-QI 61% vs post-QI 60%), the rate of hepatorrhaphy increased (pre-QI 21% to post-QI 26%; $p = 0.070$) and the OR duration increased (median pre-QI 109 minutes; interquartile range 73 to 162 minutes vs median post-QI 135 minutes; interquartile range 94 to 192 minutes; $p < 0.001$). These data might reflect more aggressive efforts to obtain hemostasis or increased stability of patients in the OR with adoption of damage control resuscitation.[20,21]

This study demonstrates how QI, research, and clinical practice can be aligned to generate evidence-based practice to improve patient care. Such agility in a trauma center is representative of the continuous improvement seen in a learning healthcare system.[1] This study serves as a model for future analyses of pre-/post-intervention studies, which are commonly used in QI. Although machine learning has been used previously to identify factors involved in decision making,[22] the use of machine learning to explain the effects of a QI intervention is novel. Our study used supervised machine learning to identify and weight factors contributing to a successful reduction in DCL. This information can be used to guide future decision making.

There are several limitations to this study. First, as with any single-institution study, the results might not be generalizable. Specifically, the safety of avoiding DCL based on early physiologic parameters or injury patterns alone should be tested in a larger multicenter trial. Surgeons have historically been taught to make the decision to perform damage control as early as possible, making trials of DCL difficult to perform. If surgeons can be convinced that the decision to perform damage control can be delayed, individual indications for DCL can be included in future trials. However, the use of machine learning to complement pre- and post-intervention studies is generalizable. Machine learning has been increasingly used to predict outcomes in trauma, such as survival, morbidity, length of stay, and hospital admission.[23] This study demonstrates a novel use of machine learning to identify factors associated with surgical decision making. Second, as often seen in clinical studies, the level of missingness was high for some variables. We attempted to address this by requiring agreement between both a complete case model and an imputed model for both time periods before designating a variable MDA or net MDA as significant. Lastly, as this was a study of a pre-/post-intervention, other variables not accounted for in this analysis might have also affected the decision to perform DCL.

CONCLUSIONS

A successful QI intervention to de-implement inappropriate DCL use in trauma patients changed surgical decision making by delaying the timing of the decision until near the end of the index operation and by clarifying the criteria for DCL. Supervised machine learning can be used in a learning trauma system to complement QI and research studies to elucidate the effects of interventions on providers' decision making.

Author Contributions

Study conception and design: Harvin, Green, Pedroza
Acquisition of data: Harvin
Analysis and interpretation of data: Harvin, Green, Pedroza, Tyson, Moore, Wade, Holcomb, Kao
Drafting of manuscript: Harvin
Critical revision: Harvin, Green, Pedroza, Tyson, Moore, Wade, Holcomb, Kao

REFERENCES

1. National Academies of Sciences, Engineering, and Medicine. A National Trauma Care System: Integrating Military and Civilian Trauma Systems to Achieve Zero Preventable Deaths after Injury. Washington, DC: The National Academies Press; 2016.
2. Deo RC. Machine learning in medicine. Circulation 2015; 132:1920−1930.
3. Rotondo MF, Schwab CW, McGonigal MD, et al. 'Damage control': an approach for improved survival in exsanguinating penetrating abdominal injury. J Trauma 1993;35:375−382; discussion 382−383.
4. Stone HH, Strom PR, Mullins RJ. Management of the major coagulopathy with onset during laparotomy. Ann Surg 1983; 197:532−535.
5. Harvin JA, Wray CJ, Steward J, et al. Control the damage: morbidity and mortality after emergent trauma laparotomy. Am J Surg 2016;212:34−39.
6. Miller RS, Morris JA Jr, Diaz JJ Jr, et al. Complications after 344 damage-control open celiotomies. J Trauma 2005;59: 1365−1371; discussion 1371−1374.
7. Higa G, Friese R, O'Keeffe T, et al. Damage control laparotomy: a vital tool once overused. J Trauma 2010;69:53−59.
8. Roberts DJ, Bobrovitz N, Zygun DA, et al. Indications for use of damage control surgery in civilian trauma patients: a content analysis and expert appropriateness rating study. Ann Surg 2016;263:1018−1027.
9. Roberts DJ, Bobrovitz N, Zygun DA, et al. Indications for use of damage control surgery and damage control interventions in civilian trauma patients: a scoping review. J Trauma Acute Care Surg 2015;78:1187−1196.
10. Harvin JA, Kao LS, Liang MK, et al. Decreasing the use of damage control laparotomy in trauma: a quality improvement project. J Am Coll Surg 2017;225:200−209.
11. Harvin JA, Wootton SH, Miller CC. Using quality improvement to promote clinical trials of emergency trauma therapies. JAMA 2018;320:1855−1856.
12. Ovretveit J, Gustafson D. Evaluation of quality improvement programmes. Qual Saf Health Care 2002;11:270−275.
13. Hartshorn S. Machine Learning with Random Forests and Decision Trees: A Visual Guide for Beginners. Amazon Digital Services; 2016.
14. Kuhn M. Applied Predictive Modeling. New York: Springer; 2013.
15. Cook D. Practical Machine Learning with H_2O: Powerful, Scalable Techniques for Deep Learning and AI. Sebastapol, CA: O'Reilly Media Inc; 2017.
16. Trickey AW, Fox EE, del Junco DJ, et al. The impact of missing trauma data on predicting massive transfusion. J Trauma Acute Surg 2013;75[Suppl 1]:S68−S74.

17. George MJ, Adams SD, McNutt MK, et al. The effect of damage control laparotomy on major abdominal complications: a matched analysis. Am J Surg 2018;216:56–59.
18. Lauerman MH, Dubose J, Cunningham K, et al. Delayed interventions and mortality in trauma damage control laparotomy. Surgery 2016;160:1568–1575.
19. Roberts DJ, Zygun DA, Faris PD, et al. Opinions of practicing surgeons on the appropriateness of published indications for use of damage control surgery in trauma patients: an international cross-sectional survey. J Am Coll Surg 2016;223:515–529.
20. Cotton BA, Reddy N, Hatch QM, et al. Damage control resuscitation is associated with a reduction in resuscitation volumes and improvement in survival in 390 damage control laparotomy patients. Ann Surg 2011;254:598–605.
21. Holcomb JB, Tilley BC, Baraniuk S, et al. Transfusion of plasma, platelets, and red blood cells in a 1:1:1 vs a 1:1:2 ratio and mortality in patients with severe trauma: the PROPPR randomized clinical trial. JAMA 2015;313:471–482.
22. Lin FP, Pokorny A, Teng C, et al. Computational prediction of multidisciplinary team decision-making for adjuvant breast cancer drug therapies: a machine learning approach. BMC Cancer 2016;16:929.
23. Liu NT, Salinas J. Machine learning for predicting outcomes in trauma. Shock 2017;48:504–510.