Evaluating the Performance of Trauma Centers: Hierarchical Modeling Should Be Used

Lynne Moore, PhD, James A. Hanley, PhD, Alexis F. Turgeon, MD, MSc, FRCPC, and André Lavoie, PhD

Background: Comparing trauma centers in terms of patient survival is a key element of performance evaluation. The current standard in trauma center profiling is based on Ordinary Logistic Regression (OLR). However, OLR does not take account of the hierarchical structure of trauma systems. Hierarchical Logistic Regression (HLR) accounts for the clustering of patients within hospitals and is therefore more theoretically appropriate. The objective of this study was to evaluate whether HLR generates different profiling results than OLR.

Methods: The study was based on the Quebec Trauma Registry with mandatory participation of all 59 designated trauma centers in the province of Quebec, uniform inclusion criteria, and standardized data collection methods. Trauma profiling was based on adjusted odds ratios, which represent the odds that a patient will die in a specific hospital compared with an “average” hospital. Risk adjustment was performed with the Trauma Risk Adjustment Model score. Hospitals were ranked according to odds ratio, and outliers were identified by comparing each hospital with all other hospitals. Hospital ranks and statistical outliers generated by OLR and HLR were compared.

Results: The study population comprised 83,504 patients including 4,731 hospital deaths (5.7%). OLR identified 11 hospitals as statistical outliers whereas HLR flagged only four of these hospitals as outliers. In addition, 54 of 59 hospitals changed ranks and 24 hospitals changed by more than five ranks when HLR replaced OLR.

Conclusions: This study shows that replacing OLR with HLR has an important impact on the results of hospital profiling. Along with the many theoretical advantages of HLR, these results support the adoption of hierarchical modeling as the standard method for trauma center profiling.

Key Words: Trauma center profiling, Patient mortality, Ordinary logistic regression, Hierarchical logistic regression, TRAM.

(J Trauma. 2010;69: 1132–1137)

Submitted for publication June 8, 2009.
Accepted for publication November 23, 2009.

Copyright © 2010 by Lippincott Williams & Wilkins
From the Department of Epidemiology and Biostatistics (L.M., J.A.H.), McGill University, Montreal; Unité de traumatologie-urgence-soins intensifs (A.F.T., A.L.), Centre de recherche du CHA and Département d’anesthésiologie (A.F.T.), Division de soins intensifs, Hôpital de l’Enfant-Jésus; and Département de médecine sociale et preventive (A.L.), Université Laval, Quebec City, Quebec, Canada.
Supported by Canadian Institutes of Health Research (Dr. Moore is the recipient of a postdoctoral research award), Canadian Health Services Research Foundation (grant RC2-1460-05), and Fonds de la Recherche en Santé du Québec (grant 015102. Dr. Tugeon is the recipient of a career research award).

Address for reprints: Lynne Moore, Unité de traumatologie-médecine d’urgence-soins intensifs, Centre de Recherche Du CHA (Hôpital de l’Enfant-Jésus), 1401, 18ème rue, Quebec City, Quebec, Canada G1J 1Z4; email: Lynne.Moore.cha@ssss.gouv.qc.ca.

DOI: 10.1097/TA.0b013e3181ce8449

MATERIALS AND METHODS

Study Population

The study was based on data from all 59 designated trauma centers in the inclusive and regionalized trauma system of the province of Quebec, Canada, collected between 1999 and 2006. During the study period, the system included 4 level I, 2 pediatric, 4 level II, 21 level III, and 28 level IV centers, designated according to American College of Surgeons guidelines.6 Data collection is mandatory for all patients meeting any of the following inclusion criteria: death after injury, intensive care unit admission, hospital stay >2 days, or transfer from another hospital. Data from individual institutions are then centralized at the ministry of health.
where they are checked for accuracy and aggregated to the Quebec Trauma Registry.

Patients dead on arrival or who delayed consultation for >48 hours after injury were excluded from this study. The identity of hospitals is concealed to respect institutional confidentiality. The study was approved by the Research Ethics Committee of the “Centre Hospitalier Affilié Universitaire de Québec” and the “Commission d’Accès à l’Information du Québec.”

Data

The QTR contains anatomic injury codes, indicators of physiologic response to injury, and patient demographics. Anatomic injury is described with Abbreviated Injury Scale codes. Physiologic response to injury is described using the Glasgow Coma Scale, systolic blood pressure, and respiratory rate, all measured on arrival in the emergency room of the treating hospital. Missing physiologic data, a common problem in trauma registries, were handled with multiple imputation, as described elsewhere. Information on comorbidities was obtained by linking the QTR to the Quebec discharge database “MED-ECHO” that is completed for all hospital admissions, using patients’ unique provincial health coverage identifier. Matching was achieved for 95% of QTR observations. The remaining observations were either patients residing outside the province of Quebec (3.8%) or patients who died in the emergency room before hospital admission (1.2%). Comorbid status was described as the number of comorbidities present before injury according to the list elaborated by Charlson et al. For the 5% patients who could not be linked to the provincial discharge database, the variable “number of comorbidities present before injury” was simulated via multiple imputation. The outcome of interest was hospital mortality.

Risk Adjustment

Adjustment for patient baseline risk was performed with the Trauma Risk Adjustment Model (TRAM), described in detail elsewhere. The TRAM is a logistic generalized additive model including indicators of anatomic injury severity (the Abbreviated Injury Scale severity score of the two most severe injuries), the body region of the most severe injury, physiologic response to injury (the Glasgow Coma Scale, systolic blood pressure, and respiratory rate), and physiologic reserve (age and the number of comorbidities). Quantitative variables are modeled with cubic splines using a fixed smoothing parameter equivalent to 4 degrees of freedom to accommodate nonlinear relations to the logit of mortality. The TRAM was fitted to study data to calculate a risk score for each patient, which is the logit of mortality conditioned on all risk factors.

Trauma Center Profiling

Standardized Mortality Ratios, used in TRISS methodology, were not used for trauma center profiling, as they are theoretically not comparable across centers and may therefore not be appropriate for ranking institutions. Trauma center performance was quantified using hospital odds ratios (OR) along with 95% confidence intervals (CI). ORs represent the odds that a patient will die in a specific hospital compared with an “average” hospital. Statistical outliers were defined as trauma centers with an OR significantly different from one (p < 0.05). Hospital ranks were generated by ordering OR estimates.

In the OLR model, trauma centers were modeled explicitly as follows:

\[
\log \frac{\pi_{ij}}{1 - \pi_{ij}} = \alpha_i + \beta_i \text{TRAM}_i + \sum_k \beta_k \text{HOSPITAL}_{ijk}
\]

where \(i\) and \(j\) indicate observations on the patient and hospital level, respectively; \(\pi\) is the probability of mortality for patient \(i\) in hospital \(j\); TRAM is the risk score generated by the TRAM; and HOSPITAL is a series of 58 (k) dummy variables representing each trauma center.

In the HLR model, trauma centers were represented by a random intercept as follows:

\[
\log \frac{\pi_{ij}}{1 - \pi_{ij}} = \alpha_i + \beta_i \text{TRAM}_i
\]

is the first (patient) level of the model, where \(i\) and \(j\) indicate observations on the patient and hospital level, respectively; \(\pi\) is the probability of mortality for patient \(i\) in hospital \(j\); and TRAM is the risk score generated by the TRAM.

\[
\alpha_i = \alpha + U_i
\]

is the second (hospital) level of the model, where the intercept is modeled as a function of a constant (\(\alpha\)) and between-hospital variance (\(U\)).

All statistical analyses were performed with the GLIMMIX procedure from the Statistical Analysis System version 9.1 (SAS Institute, Cary, North Carolina). As multiple imputation generated five imputes for each missing value, analyses were performed separately on the five datasets and pooled according to the rules described by Little and Rubin using the MIANALYZE procedure.

The agreement between statistical outliers generated by OLR and HLR was evaluated with a \([\kappa]\) coefficient. The correlation between hospital ranks was evaluated with Spearman’s correlation coefficient.

RESULTS

The study population comprised 83,504 patients of whom 33,052 (39.6%), 14,612 (17.5%), 33,105 (39.6%), and 7,466 (8.9%) were treated in a level I, level II, level III, and level IV trauma centers, respectively. A total of 78,773 (94.3%) patients survived until discharge.

OLR led to 11 statistical outliers (Fig. 1). Nine trauma centers had statistically significant higher mortality than all other centers, and two trauma centers had statistically significant lower mortality than all other centers. HLR led to four statistical outliers (Fig. 2). Two trauma centers had statistically significant higher mortality than all other centers, and two trauma centers had statistically significant lower mortality than all other centers. Note that the hospital outliers identified by HLR were a subset of those identified by OLR. The \([\kappa]\) coefficient indicated moderate agreement between OLR and HLR for identifying outliers (\([\kappa] = 0.482, 95\% \text{ CI: } 0.174–0.790\)).
OLR and HLR generated very different profiling results (Figs. 1 and 2). OR generated by OLR exhibits much more variation around the global mean (OR = 1) than those generated by HLR. In addition, unlike HLR, OLR led to very unstable OR estimates (large CIs) for low-volume centers. Figure 3 illustrates the phenomenon of shrinkage. HLR uses data from the whole sample to shrink OR estimates toward the global mean (OR = 1). Unstable estimates from low-volume centers are shrunk closer to the mean than estimates from high-volume centers.

When hospital ranks generated by OLR were compared with those generated by HLR, 54 of 59 hospitals changed...
rank and 24 hospitals changed by more than 5 ranks (Fig. 4). The Spearman’s correlation coefficient between ranks generated by OLR and HLR was 0.93 (95% CI: 0.89 – 0.96).

**DISCUSSION**

This study has demonstrated that trauma center profiling results generated by HLR differ from those generated by the current standard, OLR. HLR leads to more stable effect estimates, less statistical outliers, and different hospital ranks than OLR. Although OLR labeled nine trauma centers as worse than average performers, HLR identified only two. In addition, half of the trauma centers changed rank when OLR was used over HLR. Considering the political implications of profiling results, the observed differences are likely to be of considerable importance to physicians, hospital administrators, and Ministry of Health officials.
The difference in profiling results generated by OLR and HLR are related to the theoretical advantages of HLR. First, OLR considers the probability of mortality of each patient to be unrelated to that of all other patients, whereas HLR accounts for the fact that the outcome of patients treated in the same hospital is likely to be correlated. Thus, in HLR, two components of mortality variation are measured: variation between hospitals and variation within hospitals. This leads to an appropriate increase in the variance of OR estimates. Although it is true that this lack of independence between observations can be accommodated in OLR by using robust variance estimates, HLR models the hierarchical structure of a trauma system directly. Second, to identify outliers, each hospital is compared with an “average” hospital. This implies using multiple comparisons, which inflates the probability of falsely flagging a hospital as an outlier (type I error). Because our analysis involved 59 hospital comparisons and an [alpha] fixed at 0.05, the probability of identifying at least one outlier even if none were present was approximately 1 – (1 – 0.05)^59 = 94.8%. Unlike OLR, HLR addresses the problem of multiple comparisons because OR estimates are based on data from all hospitals rather than individual hospitals. Third, HLR provides stable OR estimates for low-volume trauma centers via shrinkage. This means that hospital OR is shrunk toward the global mean (OR = 1) by a factor that is a function of sample size; the smaller the sample size, the more OR will be shrunk toward 1. Shrinkage addresses the fact that low-volume hospitals are more likely to have extreme observed mortality rates because of random variation, a phenomenon known as regression to the mean. Efron and Morris argue that estimates derived through shrinkage are more suitable for policy making, for ordering (i.e., ranking), and for group comparisons (i.e., inter-hospital comparisons) than conventional estimates.

In acute care units, such as coronary care units and intensive care units, hierarchical regression is considered to be state of the art and is now widely implemented for profiling. HLR has been shown to produce different profiling results than OLR in these fields. Although trauma center profiling is still being performed with OLR, HLR has been used for other analysis involving trauma data.

In addition to the theoretical advantages described earlier, HLR provides a more flexible framework for trauma data analysis. Indeed, HLR can be extended to address more complex analyses of trauma systems by incorporating designation levels, geographic regions, or even individual physicians. It can also be used to evaluate time trends and to explain interhospital mortality variations by adding explanatory variables that measure, for example, structures and processes. Hierarchical regression should not be limited to profiling in trauma research. Data that are collected over trauma centers are unlikely to respect the postulate of independence in observations required for ordinary regression. Hierarchical regression allows the analyst to adjust for within-center clustering of observations present in such data.

Limitations

The objective of this study was to evaluate the impact of replacing OLR with HLR on trauma center profiling. Results should not be seen as a definitive evaluation of trauma center performance in the QTR for several reasons. First, we only had information on hospital mortality when outcomes should ideally be evaluated during a fixed period of time. Second, we did not take account of transfer status or level of care in our evaluation. Third, interhospital variations in patient mortality can be due to a multitude of factors other than quality of care such as information bias due to data quality issues, selection bias, and variations in baseline risk that remain unaccounted for.

Limitations of HLR may include its complexity and the fact that statistical methods and software to implement HLR are still under development. However, the most widely used statistical analysis software packages now include procedures dedicated to generalized linear mixed models, a family that includes HLR.

CONCLUSIONS

The theoretical advantages of HLR over OLR for hospital profiling have already been established. This study has demonstrated that replacing OLR with HLR has a real impact on the results of trauma center performance analysis. Evidence suggests that trauma profiling results generated by HLR are more valid and more reliable than those generated by OLR. We therefore recommend that HLR replace OLR as the standard analytical tool for trauma center profiling.

REFERENCES