Comparing regression-adjusted mortality to standardized mortality ratios for trauma center profiling

Lynne Moore, James A Hanley, Alexis F Turgeon¹,², André Lavoie¹

ABSTRACT

Background: Trauma center profiling is commonly performed with Standardized Mortality Ratios (SMRs). However, comparison of SMRs across trauma centers with different case mix can induce confounding leading to biased trauma center ranks. We hypothesized that Regression-Adjusted Mortality (RAM) estimates would provide a more valid measure of trauma center performance than SMRs. Objective: Compare trauma center ranks generated by RAM estimates to those generated by SMRs. Materials and Methods: The study was based on data from a provincial Trauma Registry (1999-2006; n = 88,235). SMRs were derived as the ratio of observed to expected deaths using: (1) the study population as an internal standard, (2) the US National Trauma Data Bank as an external standard. The expected death count was calculated as the sum of mortality probabilities for all patients treated in a hospital conditional on the injury severity score, the revised trauma score, and age. RAM estimates were obtained directly from a hierarchical logistic regression model. Results: Crude mortality was 5.4% and varied between 1.3% and 13.5% across the 59 trauma centers. When trauma center ranks from internal SMRs and RAM were compared, 49 out of 59 centers changed rank and six centers changed by more than five ranks. When trauma center ranks from external SMRs and RAM were compared, 55 centers changed rank and 17 changed by more than five ranks. Conclusions: The results of this study suggest that the use of SMRs to rank trauma centers in terms of mortality may be misleading. RAM estimates represent a potentially more valid method of trauma center profiling.

Key Words: Mortality, profiling, trauma

INTRODUCTION

The evaluation of trauma center performance in terms of patient outcome is a key element in the quest to improve trauma care. The most widely used indicator of trauma center outcome performance is hospital mortality, commonly described using the Standardized Mortality Ratio (SMR).¹ SMRs are calculated as the ratio of the number of deaths observed in the trauma center under evaluation to the number that would be expected if the patients were treated in a “standard” population. This standard can be based on data from the trauma system under evaluation (internal standard) or derived from national or international data (external standard). In trauma, commonly used external standards include the Major trauma Outcome Study (MTOS)² and more recently, the US National Trauma Data Bank (NTDB).³

Much attention has been paid to evaluating the adequacy of the TRISS in terms of case mix adjustment,⁴ but little attention has been paid to the use of SMRs despite their important documented limitations.⁵⁶⁷⁸ Indeed, as SMRs are adjusted according to the case mix of the hospital under evaluation, inter-hospital comparisons of performance are likely to be biased.⁴⁵⁶⁷⁸ Conversely, Regression-Adjusted Mortality (RAM) estimates, recently proposed to evaluate trauma care,⁹ are adjusted according to the global case-mix distribution of all
trauma centers, thus leading to unbiased inter-center comparisons (in the absence of other sources of bias).

We hypothesized that RAM estimates would be more appropriate than SMRs for trauma center profiling. The objective of this study was thus to evaluate whether trauma center ranks generated by RAM would differ to those generated by SMRs.

**MATERIALS AND METHODS**

The study was based on the trauma system of the province of Quebec, Canada. The trauma system in Quebec is inclusive and regionalized with 59 designated hospitals comprising urban level I trauma centers through to rural level IV community hospitals. Pre-hospital transport and transfer protocols ensure that major trauma victims are treated within the system in a center offering the appropriate level of care. Data was drawn from the Quebec Trauma Registry (1998 to 2006), which is mandatory for the 59 designated hospitals and contains information on all patients meeting any of the following inclusion criteria: death, admission to intensive care, hospital stay of more than two days, or transfer from another hospital. Registry data are extracted from patient files by medical archivists who use coding protocols that are standardized across trauma centers. The registry is centralized at the Quebec Ministry of Health and is subject to extensive validation procedures. Patients dead on arrival, patients who delayed consultation for more than 48 hours following injury, and patients with an isolated hip fracture were excluded from the study population. The study was restricted to patients with blunt trauma (94% of the study population). 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’.

Adjustment for patient case mix was performed with the Trauma Injury Severity Score (TRISS). The TRISS is derived using a logistic regression model of hospital mortality on the following three risk factors: (1) the Injury Severity Score (ISS), ranging between 1 and 75 with increasing anatomic injury severity, (2) the Revised Trauma Score (RTS), ranging between 0 and 7.48 with increasing physiological injury severity, and (3) a binary indicator for age with a cut-off at 55 years. Missing physiological data, a common problem in trauma registries, were handled with multiple imputation.

SMRs were calculated using an internal and an external standard. Internal SMRs were calculated by applying the TRISS model to study data and calculating a predicted probability of mortality for each patient. These predicted probabilities were then summed to obtain the expected number of deaths for each institution. SMRs estimates were generated by modeling observed and expected deaths counts with a random-intercept hierarchical Poisson regression model. External SMRs were obtained in the same way, except that patient-level predicted mortality probabilities were obtained by applying published TRISS coefficients from the National Trauma Data Bank to study data. The National Trauma Data Bank is the largest aggregation of trauma currently available with information on over 2 million patients treated in around 400 hospitals in the US and it is the most-widely used contemporary standard in trauma center profiling. The details of the models used are given in an Appendix.

RAM estimates were obtained directly from a hierarchical logistic regression model where TRISS components (ISS, RTS, and AGE55+) were modeled as fixed factors and trauma centers were represented by a random intercept (see Appendix). Due to the logit transformation, estimates of mortality risk generated by a logistic regression model are skewed towards zero. We therefore multiplied them by a corrective factor: the ratio of mean observed mortality to mean predicted mortality. SMR and RAM estimates were then ordered to generate hospital ranks, which were compared.

**RESULTS**

The study population comprised 83,504 patients including 4,731 hospital deaths. The lowest-volume trauma center received 34 patients while the highest-volume center received 9,632 patients over the study period. There was an important variation in the distribution of risk factors across trauma center; mean ISS varied between 5.3 and 15.8 and mean RTS varied between 6.6 and 7.4, while the proportion of patients 55 years of age of over varied between 0% (pediatric hospitals) and 76%. Crude mortality varied between 1.3% and 14.7%.

RAM estimates ranged between 2.54% (2.23%-2.90%) and 8.46% (5.17%-13.75%; [Figure 1]). Internal SMRs ranged between 0.74 (95% CI: 0.68-0.80) and 1.24 (1.02-1.52; [Figure 2]), and external SMRs ranged between 0.82 (0.76-0.90) and 1.52 (1.19-1.94; [Figure 3]).

When trauma center ranks were assigned with internal SMRs instead of RAM estimates, 49 out of 59 centers changed rank and 6 changed by more than five ranks [Figure 4]. SMRs calculated with an external standard generated ranks that were even further from those generated by RAM estimates; 55 centers changed rank and 17 changed by over 5 ranks [Figure 5]. The mean absolute difference in ranks generated by RAM and internal SMRs was 2.95 (95% CI: 2.17-3.73). The mean absolute difference in ranks generated by RAM and external SMRs was 4.27 (95% CI: 3.38-5.17).

**DISCUSSION**

In this retrospective cohort study, we observed a statistically significant difference in trauma center ranks when SMR were used over RAM estimates to evaluate the performance of trauma centers. The difference was even greater when an external standard (the NTDB) was used over an internal standard (the Quebec trauma system) to generate SMR. These results support the theoretical arguments against SMR comparisons by...
suggesting that using SMR, particularly with an external standard, may lead to biased performance evaluations. We therefore recommend that RAM estimates be used over SMR to evaluate trauma center performance in terms of mortality.

The problem of comparing SMRs was recognized as early as 1986 by Breslow and Day[1] and has been discussed frequently in the literature since then.[1,6-8,16-18] Two other important reference textbooks recommend against comparing SMR.[6,19] To circumnavigate the problem, several authors have used regression models to adjust SMR group comparisons.[20,21] Others have suggested the use of directly standardized rates (RAMs are standardized indirectly).[7,19] However, few studies have evaluated the validity of SMR comparisons empirically. Goldman et al. reported that SMR comparison was problematic only if the distribution of risk factors varied enormously between groups.[17] However, this study was performed on simulated data and with only one categorical risk factor (age). As we have illustrated, the situation is likely to be different in real data where risk adjustment often involves a series of complex risk factors.

Proponents of SMRs set forth three arguments to promote their use.[7,8,17] Firstly, SMRs can be calculated in the absence of information on the distribution of risk factors among deaths. However, this argument does not apply to trauma center profiling as information for deaths is as easily available as for...
survivors. Secondly, SMRs give precise estimates for low-volume hospitals because risk estimates are based on a large standard population. However, this second argument has largely been refuted with the advent of hierarchical regression that generates precise estimates for even the smallest hospitals through data shrinkage.\[24\] Thirdly, SMRs can be used to perform national or international comparisons. However, this argument implies that one identify a standard population that is representative of the trauma system under evaluation, which is not straightforward.\[1,8\] Indeed, the standard population should be subject to similar pre-hospital transport and transfer protocols and similar admission and discharge protocols. In addition, the database representing the standard population should have similar population coverage, data coding, and data quality. Therefore, while they are certainly of interest, valid national or international comparisons may be unrealistic. Conversely, RAM estimates can be compared across trauma centers because they are all adjusted according to the same risk factor distribution. In addition, they are more intuitive than SMRs because they directly provide information on baseline risk.

**Limitations**

This study was based on a database with good population coverage of major trauma, and standardized data collection procedures. However, several aspects may compromise the validity of profiling results. Firstly, the risk adjustment model used (TRISS) has documented limitations,\[5,23\] and alternatives have been proposed.\[29\] However, the TRISS is the only risk adjustment model available with published weights from a widely-used standard, i.e., the US National Trauma Data Bank. Furthermore, comparison of internal SMRs and RAM using the Trauma Risk Adjustment Model (TRAM) that has been shown to address the major theoretical limitations of the TRISS and to have excellent predictive validity,\[23\] led to similar results in the study population (44 hospitals changed ranks and four changed by more than 5 ranks).

Secondly, while it is the biggest aggregation of trauma data available, the US National Trauma Data Bank is based on voluntary participation, non-uniform inclusion criteria, and non-standardized data collection procedures.\[24\] In addition, admission and discharge protocols within the US trauma system are unlikely to be representative of those seen in the province of Quebec. This highlights the problem of identifying a suitable external standard. These limitations are a concern in any trauma center performance evaluation and should always be considered when interpreting profiling results. However, the objective of the present study was to compare trauma center profiling results generated by RAM estimates and SMRs rather than to offer a definitive judgment on hospital performance within the Quebec trauma system.

Finally, the use of ranks for hospital profiling has been widely criticized.\[24-26\] Indeed, ranks give no idea of the size of mortality differences between hospitals and hospitals with very different ranks may not be significantly different. However, ranks are still widely used for profiling hospitals and schools and they are unlikely to fall into misuse in the foreseeable future.\[27\] Nonetheless, we do recommend that ranks be accompanied by a measure of their uncertainty for definitive judgments of hospital performance.\[23\]

**CONCLUSIONS**

Hospital ranking is high-stakes business for clinicians, hospital administrators, policy makers as well as patients and relies on the use of valid methodology. Unlike SMRs, RAM estimates give direct information on mortality risk, do not involve the difficult choice of a suitable external standard, and are comparable over hospitals. This study demonstrates that in practice, SMRs and RAM can lead to a statistically significant difference in hospital rankings. The theoretical and empirical evidence therefore suggests that RAM estimates are more appropriate than SMRs for hospital profiling.

**APPENDIX. STATISTICAL MODELS**

\( i \) Regression Adjusted Mortality (RAM) estimates:

Random intercept hierarchical model

\[
\frac{\pi_y}{1-\pi_y} = \beta_0 + \beta_1 \text{ISS}_ij + \beta_2 \text{RTS}_ij + \beta_3 \text{AGE}_{\geq55ij}
\]

Where \( \pi \) is the probability of mortality for patient \( i \) in hospital \( j \)

\( \text{ISS}: \) Injury Severity Score; \( \text{RTS}: \) Revised Trauma Score

\( ii \) Standardized Mortality Ratios (SMR):

Trauma and Injury Severity Score (TRISS) model

\[
\frac{\pi_y}{1-\pi_y} = \beta_0 + \beta_1 \text{ISS}_ij + \beta_2 \text{RTS}_ij + \beta_3 \text{AGE}_{\geq55ij}
\]

For internal SMRs the weights, derived from fitting the TRISS to the study population, were as follows: \( \hat{\beta}_0=1.7242 , \hat{\beta}_1=0.0648, \hat{\beta}_2=-1.0287, \hat{\beta}_3=2.1733 \)

For external SMRs the weights, derived from the US National Trauma Data Bank and published for international use, were as follows: \( \hat{\beta}_0=-0.0350 , \hat{\beta}_1=0.0814, \hat{\beta}_2=-0.8142, \hat{\beta}_3=1.9623 \)

Random-intercept Poisson model

\[
\log(\mu_i) = \log(\text{E}_ij) + \beta_0 
\]

Where \( \log(\text{E}_ij) \) is an offset based on the number of deaths expected in hospital \( j \).

**REFERENCES**

3. Millham FH, LaMorte WW. Factors associated with mortality in trauma.


Received: 28.10.2011. Accepted: 04.12.2011.

Source of Support: Canadian Institutes of Health Research (Postdoctoral research award); Fonds de la Recherche en Santé du Québec (Grant number: 015102). Conflict of Interest: None declared.