Hips Can Lie: Impact of Excluding Isolated Hip Fractures on External Benchmarking of Trauma Center Performance

David Gomez, MD, Barbara Haas, MD, Mark Hemmila, MD, Michael Pasquale, MD, Sandra Goble, MS, Melanie Neal, MS, N. Clay Mann, PhD, Wayne Meredith, MD, FACS, Henry G. Cryer, MD, PhD, FACS, Shahid Shafi, MD, MPH, FACS, and Avery B. Nathens, MD, PhD, FACS

Background: Trauma centers (TCs) vary in the inclusion of patients with isolated hip fractures (IHFs) in their registries. This inconsistent case ascertainment may have significant implications on the assessment of TC performance and external benchmarking efforts.

Methods: Data were derived from the National Trauma Data Bank (2007–8.1). We included patients (aged 16 years or older) with Injury Severity Score value ≥9 who were admitted to Level I and II TCs. To ensure data quality, we limited the study to TC that routinely reported comorbidities and Abbreviated Injury Scale codes. IHF were defined as patients, aged 65 years or older, injured as a result of falls, with Abbreviated Injury Scale codes for hip fracture and without other significant injuries. TCs were stratified according to their reported inclusion of IHF in their registry. Observed-to-expected mortality ratios were used to rank TC performance first with and then, without the inclusion of patients with IHF.

Results: In total, 91,152 patients in 132 TCs were identified; 5% (n = 4,448) were IHF. The proportion of IHF per TC varied significantly, ranging from 0% to 31%. When risk-adjusted mortality was evaluated, excluding patients with IHF had significant effects: 37% (n = 49) of TCs changed their performance rank by ≥3 (range, 1–25) and 12% of centers changed their performance quintile. The greatest change in rank performance was evident in centers that routinely include IHF in their registries.

Conclusions: Given the fact that IHFs in the elderly significantly influence risk-adjusted outcomes and are variably reported by TCs, these patients should be excluded from subsequent benchmarking efforts.

Key Words: Quality improvement, Risk adjustment, Isolated hip fractures, Trauma center performance, Covariate imbalance.

(*J Trauma.* 2010;69: 1037–1041)

Submitted for publication November 23, 2009.
Accepted for publication August 9, 2010.

From the Division of Trauma (D.G., B.H., A.B.N.), Department of Surgery, Keenan Research Centre, Li Ka Shing Knowledge Institute of St. Michael's Hospital, University of Toronto, Toronto, Ontario, Canada; Department of Surgery (M.H.), University of Michigan Medical School, Ann Arbor, Michigan; Department of Surgery (M.P.), Lehigh Valley Hospital, Lehigh, Pennsylvania; National Trauma Data Bank (S.G., M.N.), American College of Surgeons, Chicago, Illinois; Intermountain Injury Control Research Center (N.C.M.), University of Utah, Salt Lake City, Utah; Department of Surgery (W.M.), Wake Forest University, Winston-Salem, North Carolina; Department of Surgery (H.G.C.), University of California Los Angeles Medical Center, Los Angeles, California; and Department of Surgery (S.S.), University of Texas Southwestern Medical School, Dallas, Texas.
Supported by Canada Research Chair Program grant (to A.B.N.).
Presented as a poster at the 23rd Annual Meeting of the Eastern Association for the Surgery of Trauma, January 19–23, 2010, Phoenix, Arizona.
Address for reprints: David Gomez, MD, 30 Bond Street, Queen Wing, 3–076, Toronto, Ontario, Canada M5B 1W8; email: gomezjaramid@smh.ca.

DOI: 10.1097/TA.0b013e3181f65387

External benchmarking of trauma center (TC) outcomes is heavily dependent on the quality and comparability of the data in hospital-based trauma registries. However, great variability exists across trauma registries with regard to the types of patients who are captured and how their characteristics are defined.[1] Recently, the introduction of the National Trauma Data Standard (NTDS) has provided registries with a uniform set of trauma registry variables, associated variable definitions, and a core set of trauma registry inclusion criteria.[2] The NTDS is likely to significantly improve the data quality for injury surveillance, outcome evaluation, and performance improvement.

The core set of trauma registry inclusion criteria described in the NTDS is designed to identify patients who sustain acute traumatic injury, and thereby capture elderly patients with isolated hip fractures (IHFs). Among patients aged 65 years or older, up to 47% of injury admissions are related to IHFs, and patients in this age group represent the fastest growing segment of the population.[3] However, elderly patients with IHFs differ significantly from the majority of patients included in trauma registries. Patients with IHF are neither typically directed to TCs through field trauma triage criteria nor seen by a multidisciplinary trauma team.[4] Furthermore, TCs vary significantly in the inclusion of IHFs in their trauma registries, which implies their care is not consistently part of TC performance improvement activities.[1,5]

External benchmarking of TC performance is accurate if there is a consistent case ascertainment across centers. Given that case ascertainment of patients with IHFs does not seem to be uniform, we postulated that this variability might have significant impact on the ability to compare outcomes across centers. We set out to evaluate the effect of this variability in registry inclusion criteria on external benchmarking of TC performance.

PATIENTS AND METHODS

Overall Study Design

We estimated risk-adjusted mortality in centers contributing to the National Trauma Data Bank (NTDB) and expressed "performance" as an observed-to-expected (O:E) mortality ratio. TCs' performances were ranked based on the O:E ratio. We then created separate risk adjustment models both including and excluding IHFs and evaluated the impact of excluding IHFs on TC performance ranking. This project was reviewed and approved by the St. Michael's Hospital Research Ethics Board.

Data Sources and Inclusion Criteria

Data were derived from the American College of Surgeons (ACS) NTDB (version 8.1). Patients (aged 16 years or older) with an Injury Severity Score (ISS) value ≥9 who were admitted at ACS-verified Level I or Level II TCs and discharged in 2007 were identified. To ensure data quality, we limited analysis to centers that routinely reported Abbreviated Injury Scale (AIS) codes and comorbidities. Patients with gunshot wounds in the head were excluded from the cohort given their high likelihood of death. Further patients with burns, asphyxiation, drowning, and overexertion as their mechanisms of injury were also excluded.

Isolated Hip Fractures

Patients with IHFs were defined as those aged 65 years or older who were injured as a result of a fall, in whom AIS codes for hip fracture were present (i.e., 851810.3—intertrochanteric femur fracture, 851812.3—neck femur fracture, and 851818.3—subtrochanteric femur fracture), and who had no injuries in any other AIS body region expect superficial injuries. Baseline patient and injury characteristics were compared across IHF and non-IHF patients. In addition, TCs specify their registry inclusion criteria on submission to the NTDB allowing the categorization of centers as either those that routinely do or do not include patients with IHF in their registry.

Evaluation of TC Performance

We used the O:E death ratios as a measure of performance. The O:E ratio has become a widely used measure of center performance and is the preferred performance measure of the ACS National Surgical Quality Improvement Program and the Trauma Quality Improvement Program.[6–8] The assessment of performance is then based on comparing a center's observed mortality with its risk-adjusted mortality. A center where the upper limit of the confidence interval of the O:E ratio is <1 is a low outlier—a good performer. If the lower limit of the 95% confidence interval is >1, the center is a high outlier—a poor performer. In addition to identifying performance outliers, the O:E ratio has been used to rank TCs based on their risk-adjusted mortality.[9]

To evaluate the relationship between the inclusion of patients with IHF and a center's risk-adjusted mortality, O:E mortality ratios were used to rank TC performance first with and then without the inclusion of patients with IHF. Centers were categorized into performance quintiles after both rankings. We then evaluated TCs' changes in rank and performance quintile after excluding patients with IHF. A change in rank of ≥3 or a change in performance quintile was considered clinically important.

Statistical Analyses

Means and standard deviations were calculated for continuous variables, and absolute and relative frequencies were measured for discrete variables. Proportions were evaluated using the χ^2 test. In all statistical analyses, $p < 0.05$ was considered significant. Separate logistic regression models both including and excluding IHF were created. In each model, the c-statistic (area under the receiver operating characteristic curve) exceeded 0.93 suggesting excellent discrimination. Simple imputation was used in case of missing values for shock and Emergency Department motor Glasgow Coma Scale. All data were analyzed using Statistical Analysis Software (version 9.1; SAS Institute, Cary, NC).

RESULTS

There were 91,152 patients meeting inclusion criteria discharged from 132 TCs with an overall mortality of 7% (n = 6,078). Overall, patients with IHFs accounted for 5% (n = 4,448) of the cohort. The proportion of patients with IHF per TC varied widely, ranging from 0% to 31% (median, 2.5%; interquartile range, 9.6%); 16% (n = 21) of centers had no patients with IHF in their trauma registry (Fig. 1).

We categorized TCs based on their reported inclusion of patients with IHF in their trauma registries. Almost two-third of centers (62%, n = 82) reported including patients with IHF in their registries. As each center defined IHFs slightly differently, two-thirds of centers (n = 34) that report routinely excluding patients with IHF had few patients with IHF. Conversely, five centers (6%) had no patients with IHF despite their indicating they do not exclude such patients, suggesting that other registry inclusion criteria were operative in their exclusion (Table 1).

We compared the characteristics of patients with IHF with those who were entered into the registries and did not meet this definition (Table 2). Patients with IHFs were more likely to be female with a greater number of comorbidities compared with other patients. They were less likely to be transferred from another center, and their rate of intensive care unit usage was one-seventh that of other included patients. Their mortality was half that of the remaining population included in trauma registries (3%, n = 118 vs. 7%, n = 5,960).

Consistent with the systematic differences in patients with IHF compared with the other populations included in trauma registries, there were significant differences in the populations of patients cared for at centers that included patients with IHF in their registry compared with those that did not (Table 3). Centers that included IHF had an older

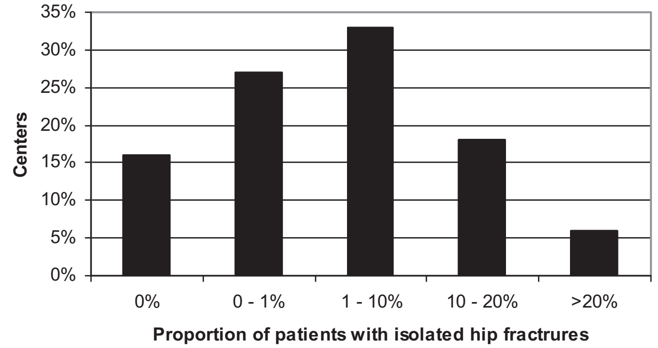

Figure 1. The proportion of patients with IHFs across Level I and II centers.

TABLE 1. Trauma Center Characteristics by IHF* Inclusion Criterion

	Trauma Centers (n = 132)	Include Patients With IHF [n = 82 (62)]	Do Not Include Patients With IHF [n = 50 (38)]	*p*
Level				0.47
I	58 (44)	38 (46)	20 (40)	
II	74 (56)	44 (54)	30 (60)	
Teaching status				0.57
Yes	54 (41)	32 (39)	22 (44)	
No	78 (59)	50 (61)	28 (56)	
Governance				0.67
For profit	6 (5)	3 (4)	3 (6)	
Nonprofit	126 (95)	79 (96)	47 (94)	

*As reported to the NTDB.
Values are presented as n (%).

TABLE 2. Baseline Characteristics of Patients With and Without IHF

	Overall (n = 91,152)	Non-IHF [n = 86,704 (95)]	IHF [n = 4,448 (5)]	*p*
Gender				<0.01
Male	60,201 (66)	58,828 (68)	1,371 (31)	
Comorbidities				<0.01
0	57,244 (63)	55,652 (64)	1,592 (36)	
1	22,570 (25)	20,938 (24)	1,632 (37)	
2	8,636 (9)	7,682 (9)	954 (21)	
≥3	2,702 (3)	2,432 (3)	270 (6)	
ICU admission	35,833 (39)	35,549 (41)	284 (6)	<0.01
Transfers	25,943 (28)	25,097 (28)	846 (19)	<0.01
Mortality	6,078 (7)	5,960 (7)	118 (3)	<0.01

ICU, intensive care unit.
Values are presented as n (%).

population, with an overrepresentation of females. Falls were more frequent in centers that included patients with IHF in their registries, and the mean ISS was lower at centers that included such patients. Significantly, lower rates of severe injuries to the head, chest, abdomen, spine, and upper extremity as evidenced by the proportion of patients with AIS value ≥3 and motor Glasgow Coma Scale score of <5 were observed in centers that included patients with IHF. Mortality was slightly but significantly higher in centers that did not include patients with IHF in their registry.

Relationship Between IHF and TC Performance

By using separate risk adjustment models, TC performance was evaluated using the O:E mortality ratio first with and then, without the inclusion of patients with IHF. Excluding patients with IHF had had marked effects on risk-adjusted mortality. More than three-quarters (78%) of centers changed their rank after patients with IHF were excluded from analyses. More than one-third of centers changed their rank by at least three (37%, n = 49). Centers that reported including patients with IHF in their registries were more likely to have rank changes of ≥3 (Fig. 2). To further evaluate the relationship between TC performance and the inclusion of patients with IHF, performance quintiles were created first with and then, without the inclusion of patients with IHF using the O:E methodology. Twelve percent (n = 16) of TCs changed their performance quintile after excluding patients with IHF.

TABLE 3. Differences in Case-Mix of Centers That Include Patients With IHF

	Overall (n = 91,152)	Include Patients With IHF (n = 59,289)	Do Not Include Patients With IHF (n = 31,863)	*p*
Gender				<0.01
Male	60,201 (66)	38,443 (64)	21,758 (68)	
Age				<0.01
16–24	17,662 (19)	10,955 (18)	6,707 (21)	
25–40	20,438 (22)	12,870 (22)	7,568 (24)	
41–54	19,454 (21)	12,316 (21)	7,138 (22)	
55–64	10,328 (11)	6,689 (11)	3,639 (11)	
65–84	18,237 (20)	12,728 (21)	5,509 (17)	
≥85	5,033 (6)	3,731 (6)	1,302 (4)	
Injury mechanism				<0.01
Motor vehicle crash	42,637 (47)	26,276 (44)	16,361 (51)	
Fall	32,498 (36)	22,734 (38)	9,764 (31)	
Firearm	4,419 (5)	2,848 (5)	1,571 (5)	
Stab	2,590 (3)	1,599 (3)	991 (3)	
Other blunt	9,008 (10)	5,832 (10)	3,176 (10)	
Injury severity by region				
Head AIS ≥3	28,286 (31)	17,613 (30)	10,673 (34)	<0.01
Chest AIS ≥3	23,660 (26)	15,112 (26)	8,548 (27)	<0.01
Abdomen AIS ≥3	5,772 (6)	3,567 (6)	2,205 (7)	<0.01
Spine AIS ≥3	2,850 (3)	1,764 (3)	1,086 (3)	<0.01
Upper extremity AIS ≥3	3,229 (4)	2,034 (3)	1,195 (4)	0.01
Lower extremity AIS ≥3	22,677 (25)	16,559 (28)	6,118 (19)	<0.01
ISS				<0.01
9–15	52,509 (58)	35,277 (59)	17,232 (54)	
16–24	23,117 (25)	14,271 (24)	8,846 (28)	
25–47	14,288 (16)	8,994 (15)	5,294 (17)	
48–75	1,238 (1)	747 (1)	491 (2)	
mGCS				<0.01
5–6	80,271 (88)	52,448 (89)	27,823 (87)	
3–4	1,921 (2)	1,246 (2)	675 (2)	
1–2	8,960 (10)	5,595 (9)	3,365 (11)	
Shock in ED				<0.01
SBP >90	86,497 (95)	56,453 (95)	30,044 (94)	
SBP >60 to ≤90	3,021 (3)	1,862 (3)	1,159 (4)	
SBP ≤60	1,634 (2)	974 (2)	660 (2)	
Transfers	25,943 (28)	16,320 (27)	9,713 (30)	<0.01
Mortality	6,078 (7)	3,735 (6)	2,343 (7)	<0.01

ED, emergency department; mGCS, motor Glasgow Coma Scale, SBP, systolic blood pressure.
Values are presented as n (%).

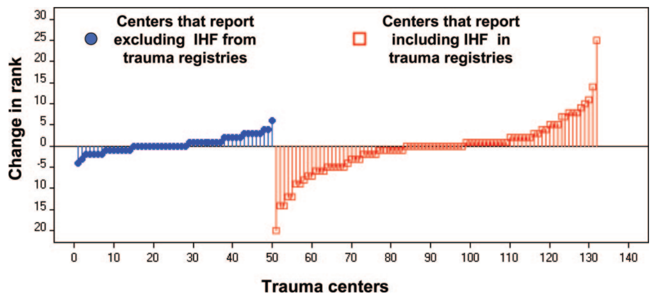

Figure 2. Change in performance rank after excluding IHF. Each center's change in performance rank after excluding patients with IHF is shown. Trauma centers were categorized based on their reported inclusion of patients with IHF in their trauma registry. A positive change in rank represents an improvement in ranking after excluding patients with IHF; conversely, a negative change in rank represents a worsening in ranking after excluding patients with IHF.

DISCUSSION

External benchmarking of TC outcomes is based on the premise that the detection of performance outliers will result in centers identifying opportunities for improvement that might not have been evident during routine center performance improvement activities.[7,8] However, external benchmarking is heavily dependent on data quality and the risk adjustment methodology used to rank performance. In an effort to improve data quality, the NTDS was developed, and trauma registries were provided with a core set of inclusion criteria and variable definitions. Nevertheless, a subgroup of patients who does not benefit from the core processes of injured patient care (e.g., field triage and trauma team activation) and has the potential to substantially influence risk-adjusted mortality is captured by the NTDS; IHFs in the elderly. Further great variability exists regarding the inclusion of IHFs in trauma registries. In a US survey that evaluated the comparability of state-wide trauma registries, 18 states reported six different exclusion criteria for same level falls comprised a mix of different age cutoffs, length of stay, and International Classification of Diseases-9 codes.[1] In addition, in a previous version of the NTDB (version 7), 17% of TCs included IHFs in those aged 65 years or older in their registries. Patients aged 65 years or older with IHFs represented between 0% and 44% of the trauma registry population across these centers.[5] We set out to evaluate the effect of this variability in registry inclusion criteria of patients with IHF on external benchmarking of TC performance.

We identified that there was considerable variability in the proportion of patients with IHF included in TC registries, ranging from as low as 0% to as high as 31% of all patients. Patients with IHF were more likely to have significant comorbidities, lower rates of intensive care unit usage, and significant lower mortality than the remaining trauma population. When the populations at centers with and without IHF were compared, the distribution of age, gender, mechanism of injury, and mean ISS differed significantly. The risk-adjusted ranking of centers compared with their peers changed markedly after patients with IHF were excluded, with more than one-third of centers changing their performance rank by ≥3 and 12% changing their performance quintile. Centers that reported including patients with IHF in their registry were those that evidenced the greatest rank changes.

The above findings are important as it is well documented that covariate imbalance leads to biased risk-adjusted outcomes.[10–12] Although, intuitively, risk adjustment models should account for systematic differences in patient characteristics across centers, most modeling approaches are dependent on centers having comparable risk profiles.[11] Regression modeling assumes that there is an overlap of risk distribution. If a group of centers systematically excludes a population of patients included at other centers, then this overlap cannot occur. Expected outcomes for the subgroup of patients only treated at one center (i.e., no overlap in risk distribution) can be extrapolated from the overlap group only by using unverifiable assumptions.[11]

The comparison of risk-adjusted outcomes when covariate imbalance is ignored is termed as the case-mix fallacy.[13] This theoretical assumption of biased estimates has been widely discussed in the context of unmeasured confounders.[11,13,14] However, IHFs in the elderly are a known and measured factor that lead to predictable covariate imbalance, particularly if we acknowledge the known variability in this particular registry inclusion criterion.

External benchmarking efforts, such as the National Surgical Quality Improvement Program, have been proven to be effective in decreasing complications and mortality.[7,15] The success of this program can be attributed in part to the focus placed on data quality and robust risk adjustment methodologies.[16,17] To obtain similar results, homogeneity of the population used during benchmarking of TC outcomes is essential to obtain unbiased risk-adjusted outcomes. The development of the NTDS was an essential first step to improve data quality. However, given the fact that IHFs in the elderly significantly influence risk-adjusted outcomes and are variably reported by TCs, these patients should be excluded from subsequent benchmarking efforts.

REFERENCES

1. Mann NC, Guice K, Cassidy L, Wright D, Koury J. Are statewide trauma registries comparable? Reaching for a national trauma dataset. *Acad Emerg Med*. 2006;13:946–953.
2. National Trauma Data Bank. National Trauma Data Standard. In: Fildes JJ, Guice K, Mann NC, Neal M, Nielsen P, Wright D, eds. *Data Dictionary, Version 1 2 2*. Chicago, IL: National Trauma Data Bank; 2008.
3. Clark DE, DeLorenzo MA, Lucas FL, Wennberg DE. Epidemiology and short-term outcomes of injured medicare patients. *J Am Geriatr Soc*. 2004;52:2023–2030.
4. Sasser SM, Hunt RC, Sullivent EE, et al., Centers for Disease Control and Prevention. Guidelines for field triage of injured patients. Recommendations of the National Expert Panel on Field Triage. *MMWR Recomm Rep*. 2009;58:1–35.
5. Committee on Trauma American College of Surgeons. National Trauma Data Bank Version 7.1. Chicago, IL: Committee on Trauma American College of Surgeons; 2007.
6. Shafi S, Nathens AB, Parks J, Cryer HM, Fildes JJ, Gentilello LM. Trauma quality improvement using risk-adjusted outcomes. *J Trauma*. 2008;64:599–604; discussion 604–596.

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. Shafi S, Nathens AB, Cryer HG, et al. The Trauma Quality Improvement Program of the American College of Surgeons Committee on Trauma. *J Am Coll Surg*. 2009;209:521–530, e521.
9. Nathens AB, Xiong W, Shafi S. Ranking of trauma center performance: the bare essentials. *J Trauma*. 2008;65:628–635.
10. Metnitz PG, Lang T, Vesely H, Valentin A, Le Gall JR. Ratios of observed to expected mortality are affected by differences in case mix and quality of care. *Intensive Care Med*. 2000;26:1466–1472.
11. Shahian DM, Normand SL. Comparison of "risk-adjusted" hospital outcomes. *Circulation*. 2008;117:1955–1963.
12. Nathanson BH, Higgins TL, Kramer AA, Copes WS, Stark M, Teres D. Subgroup mortality probability models: are they necessary for specialized intensive care units? *Crit Care Med.* 2009;37:2375–2386.
13. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet*. 2004;363:1147–1154.
14. Nicholl J. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. *J Epidemiol Community Health*. 2007;61:1010–1013.
15. Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. *Ann Surg*. 2009;250:363–376.
16. Davis CL, Pierce JR, Henderson W, et al. Assessment of the reliability of data collected for the Department of Veterans Affairs national surgical quality improvement program. *J Am Coll Surg*. 2007;204:550–560.
17. Birkmeyer JD, Shahian DM, Dimick JB, et al. Blueprint for a new American College of Surgeons: National Surgical Quality Improvement Program. *J Am Coll Surg*. 2008;207:777–782.

