Reliability of hospital readmission rates in vascular surgery

Andrew A. Gonzalez, MD, JD, MPH, a,c Micah E. Girotti, MD, MSc, a Terry Shih, MD, a Thomas W. Wakefield, MD, b and Justin B. Dimick, MD, MPH, a
Ann Arbor, Mich, and Chicago, Ill

Objective: The Center for Medicare and Medicaid Services recently began assessing financial penalties to hospitals with high readmission rates for a narrow set of medical conditions. Because these penalties will be extended to surgical conditions in the near future, we sought to determine whether readmissions are a reliable predictor of hospital performance with vascular surgery.

Methods: We examined 4 years of national Medicare claims data from 1576 hospitals on beneficiaries undergoing three common vascular procedures: open or endovascular abdominal aortic aneurysm repair (n = 81,520) or lower extremity arterial bypass (n = 57,190). First, we divided our population into two groups on the basis of operative date (2005-2006 and 2007-2008) and generated hospital risk- and reliability-adjusted readmission rates for each time period. We evaluated reliability through the use of the “test-retest” method; highly reliable measures will show little variation in rates over time. Specifically, we evaluated the year-to-year reliability of readmissions by calculating Spearman rank correlation and weighted k tests for readmission rates between the two time periods.

Results: The Spearman coefficient between 2005-2006 readmissions rankings and 2007-2008 readmissions rankings was 0.57 (P < .001) and weighted k was 0.42 (P < .001), indicating a moderate correlation. However, only 32% of the variation in hospital readmission rates in 2007-2008 was explained by readmissions during the 2 prior years. There were major reclassifications of hospital rankings between years, with 63% of hospitals migrating among performance quintiles between 2005-2006 and 2007-2008.

Conclusions: Risk-adjusted readmission rates for vascular surgery vary substantially year to year; this implies that much of the observed variation in readmission rates is either random or caused by unmeasured factors and not caused by changes in hospital quality that may be captured by administrative data. (J Vasc Surg 2014;59:1638-43.)

Because readmissions are common and costly, many policymakers view reducing readmissions as the rare opportunity to improve quality while decreasing cost. For example, new provisions under the Affordable Care Act target readmissions in an effort to reduce their $12 billion contribution to annual Medicare spending. In addition to setting reimbursement, the Act mandates public reporting of readmissions performance on the Hospital Compare website. Although these changes have initially affected select medical conditions and hip or knee surgery, they will be extended to include vascular surgery by 2015. Whether risk-adjusted readmission rates are a reliable indicator of quality remains unknown. To be a reliable indicator of quality, a measure should demonstrate stability over time. Thus, hospital rankings that are based on readmissions should vary little from year to year unless there were significant improvements in coordination of care. Prior studies have examined the temporal stability of various outcome measures. For instance, historical rates of surgical mortality can reliably predict future rates for some operations but not others. Beyond the theoretical implications, the concept of stability over time is of practical significance because the Affordable Care Act’s Hospital Readmissions Reduction Program (HRRP) will use data from 2008 through 2011 to determine pay-for-performance readmission penalties for fiscal year (FY) 2013. This study seeks to examine whether hospital readmission rates are persistent year to year or reflect random yearly variation. To investigate this question, we used data on Medicare beneficiaries undergoing abdominal aortic aneurysm repair or lower extremity arterial bypass. With the use of Spearman correlation, we ranked hospitals on the basis of 2005-2006 readmission rates and determined how well they predicted 2007-2008 ranks. Furthermore, we quantified the extent to which 2007-2008 rankings remained similar to 2005-2006 rankings.

METHODS

Databases, subjects. We used Medicare Provider Analysis and Review (MedPAR) files for beneficiaries...
between the ages of 65 and 99 years. MedPAR files contain administrative data from discharge abstracts for all Medicare beneficiaries hospitalized on a fee-for-service basis in acute care facilities. We used International Classification of Diseases, Ninth Revision (ICD-9) procedure codes to identify patients who underwent lower extremity arterial bypass, open abdominal aortic aneurysm repair (OAR), or endovascular aortic repair (EVAR) between 2005 and 2008. For OAR and EVAR, we excluded cases with an admitting diagnosis indicating aneurysm rupture. We selected these three operations, a priori because each has a high incidence of readmission (in excess of 10%). To harmonize our study population with the methodology used by the Centers for Medicare and Medicaid Services, we excluded patients who were discharged against medical advice (n = 157 patients), patients admitted as transfers from other hospitals or facilities (n = 5092), and patients who were nonelective admissions (n = 46,816). Finally, we excluded hospitals that did not perform at least one operation in each year of our study period (n = 745 hospitals).

**Statistical analysis.** We used the hospital as our unit of analysis. Our primary outcome was a composite hospital risk- and reliability-adjusted 30-day readmission rate. We calculated point estimates and confidence intervals for this outcome by a three-step process.

First, we used standard logistic regression to calculate an empirically derived risk score. This score was calculated through the use of a continuous variable for age, a dichotomous variable for sex, and “dummy” variables for pre-existing medical comorbidities on the basis of Elixhauser’s methods. For each hospital, a separate risk score was created for each operation and each year (12 in all). Beyond simplifying the model, a single risk score reduces the likelihood of model nonconvergence. In contrast, the use of individual variables often results in failure of the statistical program to fit an adaptive model to the covariate matrix. This approach has previously been applied to similar data in surgical patients and to the reporting of vascular outcomes. Specifically, this risk score was generated by means of the logistic postestimation command that predicts the log(odds) of readmission for each patient (-xb option in STATA). We used modeled readmissions in units of log(odds) rather than predicted probability because log(odds) is linear with respect to binary outcomes (eg, readmission).

Second, we used a hierarchical mixed-effects logistic regression model that was based on the patient risk score (described above) with a hospital variable as the random effect. The Centers for Medicare and Medicaid Services uses hierarchical logistic regression to calculate hospital readmission rates for the HRRP. Hierarchical regression is an advanced statistical technique that leverages empirical Bayes theorem to directly model variation at the hospital level. As a result, hierarchical modeling yields more stable estimates of both coefficients and standard errors. Prior studies of outcomes after vascular surgery have validated this technique and shown it to reduce statistical noise and provide more robust estimates of hospital quality compared with risk adjustment alone.

Third, we generated a composite readmission rate for 2005-2006 and for 2007-2008. For each hospital, we multiplied its procedure-specific readmission rate by its volume for that procedure; summed the resulting products; and divided the result by the hospital’s total volume for all three procedures. This adjusted for differences in volume over time and for differences in the proportion of each operation performed at any individual hospitals. Thus, a hospital that performed many bypasses but few EVARs would have a composite readmission rate closer to its bypass readmission rate.

Finally, we ranked hospitals according to risk- and reliability-adjusted 30-day readmission rates during the historical time period (2005-2006) and during the outcome time period (2007-2008). These ranks were used to stratify hospitals into quintiles of readmission for each time period. We used a reclassification matrix to analyze hospital migration among quintiles between the two time periods. Correlation was assessed by means of Spearman rank coefficients and weighted k tests. In addition to composite rankings that were based on all operations, we evaluated each operation separately.

**Sensitivity analysis.** To ensure that our results were robust to any anomalies unique to the combination of years forming the historical and outcome time periods, we analyzed reliability and reclassification across three 1-year periods: 2005-2006, 2006-2007, and 2007-2008.

**RESULTS**

We evaluated 1576 hospitals performing OAR, EVAR, and lower extremity arterial bypass between 2005 and 2008. The mean adjusted readmission rate was 19.3% (standard deviation, 2.8%; range, 15.5%-23.4%) in 2005-2006 and 15.8% (standard deviation, 1.9%; range, 13.3%-18.6%) in 2007-2008. Overall, 129 hospitals (8.2%) saw a decrease in their readmission rate by at least 5 percentage points from 2005-2006 to 2007-2008. In contrast, only 45 hospitals (2.9%) saw any increase in their readmission rate.

**Table.** Patient characteristics by cohort

<table>
<thead>
<tr>
<th>Variable</th>
<th>2005-2006</th>
<th>2007-2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal aortic aneurysm repair</td>
<td>33,973</td>
<td>47,547</td>
</tr>
<tr>
<td>Median age (IQR), years</td>
<td>76 (71-80)</td>
<td>75 (70-81)</td>
</tr>
<tr>
<td>Sex (% women)</td>
<td>24.7</td>
<td>23.9</td>
</tr>
<tr>
<td>% Black race</td>
<td>3.8</td>
<td>3.7</td>
</tr>
<tr>
<td>% 3+ Comorbidities</td>
<td>32</td>
<td>39</td>
</tr>
<tr>
<td>% Endovascular</td>
<td>59</td>
<td>69</td>
</tr>
<tr>
<td>Lower extremity arterial bypass</td>
<td>24,606</td>
<td>32,584</td>
</tr>
<tr>
<td>Median age (IQR), years</td>
<td>75 (70-81)</td>
<td>75 (70-81)</td>
</tr>
<tr>
<td>Sex (% women)</td>
<td>45.4</td>
<td>42.8</td>
</tr>
<tr>
<td>% Black race</td>
<td>12.6</td>
<td>11.5</td>
</tr>
<tr>
<td>% 3+ Comorbidities</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>% Claudication</td>
<td>27</td>
<td>30</td>
</tr>
<tr>
<td>% Tissue loss</td>
<td>43</td>
<td>43</td>
</tr>
</tbody>
</table>

ICD-9, International Classification of Diseases, Ninth Revision; IQR, interquartile range.

aStatistically significant at P < .05.

bICD-9 code 440.21 (claudication) in any diagnostic code field.

cICD-9 codes 440.22 (rest pain); 440.23 (ulceration); 440.24 (gangrene).
In evaluating patient characteristics for both year groups, there were statistically significant differences for all demographic factors with the exception of age (Table). However, the use of EVAR was the only clinically meaningful difference between 2005-2006 and 2007-2008 (30% vs 37%; \( P < .001 \)).

The Spearman coefficient for the correlation between 2005-2006 ranks and 2007-2008 ranks was 0.57 (\( P \) value for independence of ranks <.001; weighted \( \kappa = 0.42 \) (\( P \) value for agreement <.001; Fig 1). This indicates a moderate correlation between hospital rankings that were based on 2005-2006 readmission rates and rankings that were based on 2007-2008 rates. Only 32% of the variation in 2007-2008 readmission rates could be explained by hospital performance in 2005-2006 (historical time period). Restated, 68% of the variation was explained by factors other than historical performance. Thus, we observed significant fluctuation in quintile rankings between the two time periods (Fig 2). Overall, 61% of hospitals were reclassified into different quintiles of performance between 2005-2006 and 2007-2008. With the exception of hospitals in the worst quintile (highest readmission rates) for 2005-2007, there are no clinically meaningful differences between the other 80% of hospitals (Fig 3). Our results were similar when examining operations individually (data not included).

Sensitivity analysis. When analyzing changes for each year individually, we found that \( R^2 \) varied from 0.25 (2006 vs 2007) to 0.41 (2007 vs 2008).

DISCUSSION

In this study, we assessed the year-to-year variability of hospital readmission rates after major vascular surgery. To draw conclusions about groups (eg, hospitals), it is generally accepted that a measure should have reliability of at least .70.\(^4\) We found that readmission rates have suboptimal predictive reliability (.32). This implies that readmissions following major vascular surgery do not reflect a fixed set of attributes capturing a hospital’s ability to provide coordinated transitions of care (ie, optimal discharge practices). From a practical standpoint, this study has important policy implications, challenging whether Medicare’s recent implementation of penalties under the HRRP is going to incentivize hospitals in a productive way. If the penalized hospitals do not have control over the outcome (ie, readmission is due to chance alone), then assessing such penalties will not have the desired effect of reducing readmission rates.

Our main finding, that readmission rates have suboptimal reliability, implies that factors other than quality drive readmissions performance. Aside from actual quality improvement (or degradation), the changes in readmission...
rankings that we observed may be explained by chance year-to-year variation or by unmeasured patient factors. Although the problem of chance variation in outcome measures was recognized in the 1950s, prior literature around reliability remains sparse. The few existing studies evaluating the reliability of surgical outcomes have focused exclusively on postoperative mortality. For instance, in the case of coronary artery bypass grafting, elective repair of abdominal aortic aneurysm, and pancreatectomy, historical mortality predicts only one-third to one-half of the variation in subsequent mortality. The remaining variation (the majority) is due to chance or statistical “noise.”

Akin to our readmission findings, chance variation in mortality rates compared with other hospitals. Yet, as a cross-sectional measure, mortality rate may be fairly reliable, depending on the clinical context. For example, in postcolectomy patients, the reliability of mortality ranges from 0.24 to 0.45 (5th to 95th percentile of volume), but, in trauma patients, it ranges from 0.69 to 0.97 (depending on volume).

Independent of clinical context, readmissions simply may be a less reliable measure of hospital quality. Although no prior studies have examined the reliability of postoperative readmission rates, there is previous evidence from Medicare patients hospitalized for pneumonia, myocardial infarction, or heart failure. Press et al found that hospital readmission rankings fluctuated significantly during a 3-year period. In fact, hospitals with the lowest readmission rates in 2009 tended to have the highest increases in readmissions by 2011. The authors attributed their findings primarily to regression to the mean. In other words, extreme outliers observed in any one time period will migrate toward average performance as time goes on.

Our study adds to the literature by examining the reliability of readmissions in the context of vascular surgery. It is especially timely because the Centers for Medicare and Medicaid Services (CMS) has specifically targeted vascular procedures for expansion of the HRRP. After the FY 2013 evaluation, more than 2200 hospitals were penalized, with a total of $500 million. In FY 2013, only medical procedures were targeted and penalties were capped at 1% of the hospital’s operating revenue. However, by FY 2015, the HRRP will expand to include vascular procedures and the maximum cap will triple to 3%, raising aggregate potential penalties to $1 billion.

**Policy implications.** Our findings suggest that readmissions are a suboptimal proxy for hospital quality. In fact, penalizing hospitals for higher-than-expected readmission rates carries the potential for significant harm. First, it inappropriately labels some hospitals as poor performers. In turn, these institutions may divert resources toward addressing a misperceived readmissions problem that is in fact an artifact of random yearly variation. Alternatively, other hospitals with true coordination of care problems

<table>
<thead>
<tr>
<th>Major Vascular Operations</th>
<th>Ranking Based on 2007-08 Readmission Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>47% (149)</td>
</tr>
<tr>
<td></td>
<td>25% (80)</td>
</tr>
<tr>
<td></td>
<td>15% (47)</td>
</tr>
<tr>
<td></td>
<td>9% (27)</td>
</tr>
<tr>
<td></td>
<td>4% (13)</td>
</tr>
</tbody>
</table>

Fig 2. Comparison of quintile ranking between 2005-2006 and 2007-2008 for 1576 hospitals. Shading intensity represents the magnitude reclassification; (n) represents the number of hospitals in each quintile combination. Overall, 63% of hospitals were reclassified between the 2-year groups.

![Fig 3. Risk- and reliability-adjusted hospital readmission rates in 2007-2008, according to quintiles of historical readmission performance in 2005-2006. Both historical and subsequent readmission rates are adjusted for patient characteristics.](image-url)
(not captured by readmission data) may be lulled into a false sense of security.

One solution for the issues arising from unreliable readmission rates may be to directly measure coordination of care. Although coordination of care is frequently cited as essential to improving postoperative outcomes, hospitals are infrequently assessed along this domain of quality. However, there are validated instruments specifically designed to measure coordination of care. For instance, the National Quality Forum, a leading patient safety organization, endorses the Three-Item Care Transition Measure (CTM-3). The CTM-3 evaluates three domains of care coordination: patient understanding of postdischarge self-care, medication management, and use of a patient-centered discharge care plan. Whereas other instruments focus on questions asked of patients during the discharge process, the CTM-3 focuses on actions taken by providers in preparation for discharge. The CTM-3 has been shown to discriminate between patients who did versus did not have a subsequent readmission.

Moreover, CMS’s current methodology for quantifying performance under HRRP creates several unintended but perverse economic consequences. Chiefly, hospitals caring for poor and minority patients are most heavily penalized. In one study examining the impact of HRRP on safety-net hospitals, the authors found that safety-net hospitals were more than twice as likely to be penalized as other hospitals. In fact, only 20% of safety-net hospitals avoided penalties under HRRP. This situation is particularly problematic in surgery, given that Medicaid and self-insured patients often generate net revenue losses for hospitals (on a per-operation basis). Thus, HRRP may inadvertently reduce scarce resources available for quality improvement at institutions already under the most financial strain.

Lack of reliability makes the use of readmissions for setting monetary fines problematic. Yet, this does not mean that longitudinally tracking readmissions is without merit. Readmissions may nonetheless play other roles in improving the quality of care. Though not suitable as a penalty measure, readmissions may be useful as an improvement measure.

Limitations. Our study is not without limitations. First, the findings may not be widely generalizable, given that our study population included only Medicare patients ages 65 to 99 years. However, the significant overlap in age range between patients with vascular disease and the fee-for-service Medicare population somewhat mitigates this limitation. Second, our risk-adjustment model may be biased by unobserved differences in patient factors. However, the focus of the present study was to assess reliability, an attribute relatively unaffected by an imperfect ability to adjust for patient severity.

Furthermore, it is important to consider that we found 30-day all-cause postoperative readmissions an unreliable measure of quality at the hospital level. However, at the patient level, readmissions may be predictable (or even preventable), depending on the clinical context.

CONCLUSIONS

In this study, we examined the predictive reliability of readmission rates in vascular surgery. Our major finding, hospital rankings that are based on readmission rates vary from year to year due to chance (or unmeasured variables), suggests that measuring readmissions with the use of administrative data is an unreliable indicator of hospital quality. Although reducing unnecessary rehospitalization is a laudable goal, the lack of reliability raises concerns about the use of this measure for public reporting of hospital quality. From a practical standpoint, the use of an unreliable measure to financially penalize hospitals (eg, under CMS’s HRRP) has at least three unintended consequences. First, hospitals misclassified as low performers may allocate scarce resources toward correcting misperceived readmission problems. Second, the problems with coordination of care problems may remain undetected at hospitals misclassified as good performers. Finally, the policy creates perverse economic incentives that penalize the nation’s most vulnerable hospitals.

AUTHOR CONTRIBUTIONS

Conception and design: AG, MG, TS, JD
Analysis and interpretation: AG, MG, TS, JD
Data collection: AG, JD
Writing the article: AG, TS, JD
Critical revision of the article: AG, TW, JD
Final approval of the article: AG, TW, JD
Statistical analysis: AG, JD
Obtained funding: TW, JD
Overall responsibility: AG

REFERENCES
