Comparison of 30-Day Mortality Models for Profiling Hospital Performance in Acute Ischemic Stroke With vs Without Adjustment for Stroke Severity

Gregg C. Fonarow, MD
Wenqin Pan, PhD
Jeffrey L. Saver, MD
Eric E. Smith, MD, MPH
Mathew J. Reeves, PhD
Joseph P. Broderick, MD
Dawn O. Kleindorfer, MD
Ralph L. Sacco, MD
DaiWai M. Olson, PhD
Adrian F. Hernandez, MD, MHS
Eric D. Peterson, MD, MPH
Lee H. Schwamm, MD

Context There is increasing interest in reporting risk-standardized outcomes for Medicare beneficiaries hospitalized with acute ischemic stroke, but whether it is necessary to include adjustment for initial stroke severity has not been well studied.

Objective To evaluate the degree to which hospital outcome ratings and potential eligibility for financial incentives are altered after including initial stroke severity in a claims-based risk model for hospital 30-day mortality for acute ischemic stroke.

Design, Setting, and Patients Data were analyzed from 782 Get With The Guidelines–Stroke participating hospitals on 127,950 fee-for-service Medicare beneficiaries with ischemic stroke who had a score documented for the National Institutes of Health Stroke Scale (NIHSS, a 15-item neurological examination scale with scores from 0 to 42, with higher scores indicating more severe stroke) between April 2003 and December 2009. Performance of claims-based hospital mortality risk models with and without inclusion of NIHSS scores for 30-day mortality was evaluated and hospital rankings from both models were compared.

Main Outcomes Measures Model discrimination, hospital 30-day mortality outcome rankings, and value-based purchasing financial incentive categories.

Results Across the study population, the mean (SD) NIHSS score was 8.23 (8.11) (median, 5; interquartile range, 2–12). There were 18,186 deaths (14.5%) within the first 30 days, including 7,430 deaths (5.8%) during the index hospitalization. The hospital mortality model with NIHSS scores had significantly better discrimination than the model without (C statistic, 0.864; 95% CI, 0.861–0.867, vs 0.772; 95% CI, 0.769–0.776; P < .001). Among hospitals ranked in the top 20% or bottom 20% of performers by the claims model without NIHSS scores, 26.3% were ranked differently by the model with NIHSS scores. Of hospitals initially classified as having “worse than expected” mortality, 57.7% were reclassified to “as expected” by the model with NIHSS scores. The net reclassification improvement (93.1%; 95% CI, 91.6%–94.6%; P < .001) and integrated discrimination improvement (15.0%; 95% CI, 14.6%–15.3%; P < .001) indexes both demonstrated significant enhancement of model performance after the addition of NIHSS. Explained variance and model calibration was also improved with the addition of NIHSS scores.

Conclusion Adding stroke severity as measured by the NIHSS to a hospital 30-day risk model based on claims data for Medicare beneficiaries with acute ischemic stroke was associated with considerably improved model discrimination and change in mortality performance rankings for a substantial portion of hospitals.
creasing interest in also reporting outcomes for Medicare beneficiaries hospitalized with acute ischemic stroke.9,10

Risk adjustment for case mix is considered essential for accurately assessing and reporting hospital-level outcomes.3-5,11 The risk-adjustment models currently used by CMS incorporate data exclusively from administrative claims.3,5 Although claims data risk models for acute myocardial infarction, heart failure, and community-acquired pneumonia were validated against clinical data,3,5 adequate case-mix adjustment for acute ischemic stroke by claims data may be particularly difficult, and current models do not include adjustment for stroke severity. It has been previously shown in patient-level analyses that stroke severity as indexed by the National Institutes of Health Stroke Scale (NIHSS) is an important predictor of mortality in acute ischemic stroke.12-15 However, whether adjustment for stroke severity is necessary for hospital-level risk profiling has not been well studied.

Using data from hospitals participating in the Get With The Guidelines-Stroke (GWTG-Stroke) program on acute ischemic stroke admissions linked to Medicare data, this study was designed to (1) evaluate the change in model performance by adding or not adding NIHSS score for predicting hospital-level 30-day all-cause mortality for Medicare beneficiaries with acute ischemic stroke and (2) determine whether there are meaningful differences in hospital ranking with the use of models with and without adjustment for NIHSS score.

METHODS

Clinical data including stroke severity were obtained from GWTG-Stroke, and administrative claims data were obtained from CMS. Data from the GWTG-Stroke registry were linked with enrollment files and inpatient claims from CMS for the period April 1, 2003, through December 31, 2009. Follow-up continued through 2010. The design, inclusion criteria, and data collection methods for GWTG-Stroke have been described previously.16,17 Patients were eligible for inclusion in the GWTG-Stroke registry if they were admitted for acute stroke. Trained hospital personnel ascertained acute ischemic stroke admissions by either prospective clinical identification, retrospective identification using discharge codes from the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), or a combination. Patient data abstracted by trained hospital personnel included demographics, medical history, in-hospital treatment and events, discharge treatment and counseling, mortality, and discharge destination. Admission staff, medical staff, or both recorded race/ethnicity, usually as the patient was registered. Prior studies have suggested differences in outcomes based on race/ethnicity.

All patient data were deidentified before submission. All states and regions of the United States were represented, and a variety of centers participated, from community hospitals to large tertiary centers. Data on hospital-level characteristics were obtained from the American Hospital Association.18 All participating institutions were required to comply with local regulatory and privacy guidelines and, as determined by each participating institution, to obtain institutional review board approval. Outcome Sciences served as the registry coordinating center. The Duke Clinical Research Institute served as the data analysis center and has an agreement to analyze the aggregate deidentified data for research purposes. The institutional review board of the Duke University Health System approved the study.

The CMS files (100% Medicare Research Identifiable Files) included data for all fee-for-service Medicare beneficiaries aged 65 years or older who were hospitalized with a diagnosis of acute stroke (ICD-9-CM codes 430.x, 431.x, 433.x, 434.x, and 436.x). We linked patient data in the GWTG-Stroke registry with Medicare Part A inpatient claims, matching by admission and discharge dates, hospital, date of birth, and sex using methods previously described.15,17,19,20 Patients in Medicare managed care plans (15%-25% of the population depending on the region of the country) or other types of insurance are not included in fee-for-service Medicare claims files and therefore cannot be matched.15,17,19,20 Patients from centers with fewer than 25 ischemic stroke patients with NIHSS score documented during the study period were excluded to minimize the likelihood of sampling error.

NIHSS

The NIHSS is a 15-item neurologic examination stroke scale used to provide a quantitative measure of stroke-related neurologic deficit by evaluating the effect of acute ischemic stroke on the levels of consciousness, language, neglect, visual-field loss, extraocular movement, motor strength, ataxia, dysarthria, and sensory loss.15,21 The NIHSS is designed to be a simple, valid, and reliable tool that can be administered at the bedside consistently by physicians, nurses, or therapists. Each item is scored with 3 to 5 grades, with 0 as normal and the final total score having a potential range of 0 to 42, with higher scores indicating greater stroke severity. The first recorded NIHSS score, as close to admission time as possible, was collected.

Hospital-Level Outcome

The outcome of interest was hospital-level all-cause mortality within 30 days from time of admission. Deaths and dates of death were obtained, with complete ascertainment, from the CMS vital status files.15,17,19,20

30-Day Mortality Model Derivation

The approach to risk model development without the NIHSS score was to closely follow the previously described approach used by the Yale New Haven Health System/Center for Outcomes Research and Evaluation and CMS for other publicly reported 30-day mortality1-6 and described in—
veloping a model for 30-day risk-standardized mortality for acute ischemic stroke. This approach to risk adjustment generally follows the principles stated in the American Heart Association Scientific Statement “Standards for statistical models used for public reporting of health outcomes.”

The candidate variables for the risk model are patient-level risk adjustors that are expected to be predictive of mortality, based on empirical analysis, prior literature, and clinical judgment, including demographic factors (age, sex) and indicators of comorbidity. For each patient, the candidate variables considered for this model were derived from the Medicare claims files and included secondary diagnosis and procedure codes from the index hospitalization and from the principal and secondary diagnosis codes from hospitalizations, institutional outpatient visits, and physician encounters in the 12 months before the index hospitalization. The model is intended to adjust for case differences based on the diverse aspects of the clinical status of acute ischemic stroke patients at time of admission. Condition categories are drawn from more than 15,000 ICD-9-CM diagnosis codes.

The final set of risk-adjustment variables for the claims-based model was selected to be aligned with those included in the proposed CMS acute ischemic stroke 30-day mortality measure. The final 87 variables are shown in eTable 1 (available at http://www.jama.com). The approach to model development with the NIHSS score included was methodologically identical, except for the addition of a measure of disease severity (NIHSS). The NIHSS score was treated as a continuous parameter. Hierarchical generalized logistic regression models were used to model the binary outcome of mortality within 30 days of admission as a function of patient demographic and clinical characteristics and a random hospital-specific effect. This strategy accounts for within-hospital correlation of the observed outcomes. Additional details are provided in the eMethods.

### Statistical Analysis

Discrimination of the base claims model without NIHSS score was assessed by determining the C statistic and was compared with the discrimination of the claims model with NIHSS score. We used the integrated discrimination improvement (IDI) index to measure how the model that included NIHSS score reclassified patients compared with the model without NIHSS score. A higher IDI index indicates a greater improvement in risk discrimination and improved reclassification. The net reclassification improvement (NRI) index, which compares the shifts in reclassified categories by observed outcome, resulting from the addition of NIHSS score to the model was also determined. A higher NRI index indicates a greater improvement in risk discrimination and improved reclassification.

We also ranked hospitals by their 30-day adjusted mortality from each model and plotted the agreement between these rankings. We ranked these intercepts to group hospitals into 3 categories, top 20%, middle 60%, and bottom 20%, and compared the results across models. We chose these categories because they reflect categories that may be relevant to pay-for-performance programs, in which the top 20% of hospitals are eligible for bonus payments and the bottom 20% of hospitals may be subject to a payment penalty. We also grouped hospitals into the top 5%, middle 90%, and bottom 5% and compared the results across models. As an additional approach, hospitals with the 95% credible intervals of the estimated random intercepts not covering the null point are considered to have performance that is significantly better or worse than the average hospital. These categories are analogous to the portions of hospitals identified in Hospital Compare as having better than, no different than, or worse than expected 30-day risk-standardized mortality rates. All P values are 2-sided, with P < .05 considered statistically significant. SAS software version 9.2 (SAS Institute) was used for all analyses.

### RESULTS

There were 693,458 hospitalizations of patients with acute ischemic stroke enrolled in GWTG-Stroke between April 2003 and December 2009. From 472,443 hospitalizations of patients aged 65 years or older, we matched 318,393 patients (67.4%) to fee-for-service Medicare claims from 1428 hospitals. Of these acute ischemic stroke patients, there were 143,481 (45.1%) with NIHSS score documented. We further confined the population to patients’ first index stroke admission during the study period (n = 138,314), nontransferred patients with acute ischemic stroke (n = 131,404), and patients from centers with fewer than 25 ischemic stroke patients with NIHSS score documented. This resulted in a final study population of 127,950 acute ischemic stroke patients from 782 GWTG-Stroke hospitals. Among these hospitals, there were 124,428 nontransferred acute ischemic stroke patients without NIHSS score documented during the study period. The demographics, clinical characteristics, geographic distribution, and hospital characteristics of the study patients with and without NIHSS score documented from the 782 hospitals were similar, with few exceptions (eTable 2).

The characteristics of the 127,950 acute ischemic stroke patients with NIHSS score documented are shown in Table 1. The median age was 80 years, 57.3% were women, and 86.2% were white. Prior stroke or transient ischemic attack was present in 32.4% of patients. Comorbidities were common with hypertension in 82.6%, diabetes in 29.2%, coronary artery disease or prior myocardial infarction in 33.7%, and history of atrial fibrillation or flutter in 26.9%. The NIHSS median score in this overall population was 5 (interquartile range [IQR], 2-12). The median hospital-level NIHSS score was 5 (IQR, 4-7). Of the 782 GWTG-Stroke hospitals included in this study, median bed size was 377, all regions of the United States were represented, and teaching hospitals accounted for 21.3% of all hospitals and 28.1% of admis-
There were 18,186 deaths (14.5%) within the first 30 days, including 7,430 deaths during the index hospitalization (in-hospital mortality, 5.8%). The median hospital-level 30-day mortality rate was 14.5% (IQR, 11.3%-17.9%).

Table 2 reports the performance of the claims models without NIHSS score vs the claims model with NIHSS score for 30-day mortality among acute ischemic stroke patients at all registry hospitals in the analysis. Discrimination, calibration, and explained variance were substantially improved with the addition of NIHSS score. The hospital claims mortality model without NIHSS score had a C statistic of 0.772 (95% CI, 0.769-0.776), whereas the NIHSS score alone had a C statistic of 0.822 (95% CI, 0.819-0.825) and the claims model with the NIHSS score included had a C statistic of 0.864 (95% CI, 0.861-0.867; absolute difference for claims model with vs without NIHSS score, 0.091; 95% CI, 0.088-0.094; P < .001). Explained variance improved over the model without NIHSS score (Table 2).

Additional tests for model discrimination were improved with the addition of NIHSS score. The NRI index (93.1%; 95% CI, 91.6%-94.6%; P < .001) and IDI index scores (discrimination slope for model with vs without NIHSS score, 27.7% vs 12.8%; difference, 15.0%; 95% CI, 14.6%-15.3%; P < .001; relative IDI, 1.17) all demonstrated substantially more accurate classification of hospital 30-day mortality after the addition of NIHSS score to the claims model. Although both claims models resulted in wide variance of predicted risk between the most extreme deciles, the model with NIHSS exhibited better agreement between observed and predicted mortality rates (Hosmer-Lemeshow goodness-of-fit test χ², 352.9 vs 173.8; P < .001).

Of the 782 hospitals in the analysis, the median absolute change in rank position was 79 places (IQR, 35-135) when hospitals were ranked with risk models without and with NIHSS score. The numbers of hospitals in which the

### Table 1. Patient and Hospital Characteristics of the Study Cohort of Medicare Beneficiaries With Acute Ischemic Stroke

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Study Cohort (n = 127,950)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (IQR), y</td>
<td>80 (73-86)</td>
</tr>
<tr>
<td>Male sex, %</td>
<td>42.7</td>
</tr>
<tr>
<td>Race/ethnicity, %</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>86.2</td>
</tr>
<tr>
<td>Black</td>
<td>9.5</td>
</tr>
<tr>
<td>Hispanic</td>
<td>9.5</td>
</tr>
<tr>
<td>Arriveal mode, %</td>
<td></td>
</tr>
<tr>
<td>Emergency medical services</td>
<td>66.2</td>
</tr>
<tr>
<td>Private transport</td>
<td>31.8</td>
</tr>
<tr>
<td>Off-hours arrival time, %</td>
<td>50.9</td>
</tr>
<tr>
<td>Atrial fibrillation/flutter, %</td>
<td>26.9</td>
</tr>
<tr>
<td>Previous stroke/TIA, %</td>
<td>32.4</td>
</tr>
<tr>
<td>CAD/prior MI, %</td>
<td>33.7</td>
</tr>
<tr>
<td>Diabetes mellitus, %</td>
<td>29.2</td>
</tr>
<tr>
<td>Peripheral vascular disease, %</td>
<td>5.7</td>
</tr>
<tr>
<td>Hypertension, %</td>
<td>82.6</td>
</tr>
<tr>
<td>Smoker, %</td>
<td>10.9</td>
</tr>
<tr>
<td>Dyslipidemia, %</td>
<td>42.1</td>
</tr>
<tr>
<td>Body mass index, median (IQR)</td>
<td>25.8 (22.7-29.6)</td>
</tr>
<tr>
<td>NIHSS total score, mean (SD)</td>
<td>8.23 (8.11)</td>
</tr>
<tr>
<td>Median score (IQR)</td>
<td>5 (2-12)</td>
</tr>
<tr>
<td>Discharge status, %</td>
<td></td>
</tr>
<tr>
<td>Home</td>
<td>36.4</td>
</tr>
<tr>
<td>Skilled nursing facility</td>
<td>24.7</td>
</tr>
<tr>
<td>Rehabilitation</td>
<td>26.0</td>
</tr>
<tr>
<td>Transfer to acute care facility</td>
<td>1.7</td>
</tr>
<tr>
<td>Left against medical advice</td>
<td>0.2</td>
</tr>
<tr>
<td>Hospice</td>
<td>5.1</td>
</tr>
<tr>
<td>Died in hospital</td>
<td>5.8</td>
</tr>
<tr>
<td>Ambulatory status%</td>
<td></td>
</tr>
<tr>
<td>Ambulated independently</td>
<td>41.3</td>
</tr>
<tr>
<td>With assistance</td>
<td>35.7</td>
</tr>
<tr>
<td>Unable to ambulate</td>
<td>21.6</td>
</tr>
<tr>
<td>30-day mortality</td>
<td>14.5</td>
</tr>
<tr>
<td>30-day rehospitalization</td>
<td>12.0</td>
</tr>
<tr>
<td>Hospital characteristics</td>
<td></td>
</tr>
<tr>
<td>No. of beds, median (IQR)</td>
<td>377 (265-561)</td>
</tr>
<tr>
<td>No. of stroke discharges, 2009a</td>
<td></td>
</tr>
<tr>
<td>0-100</td>
<td>21.1 (299)</td>
</tr>
<tr>
<td>101-300</td>
<td>65.0 (424)</td>
</tr>
<tr>
<td>&gt;300</td>
<td>12.8 (44)</td>
</tr>
<tr>
<td>Geographic region%</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>29.1 (225)</td>
</tr>
<tr>
<td>Midwest</td>
<td>13.6 (113)</td>
</tr>
<tr>
<td>South</td>
<td>42.9 (321)</td>
</tr>
<tr>
<td>West</td>
<td>14.5 (110)</td>
</tr>
<tr>
<td>Hospital type%</td>
<td></td>
</tr>
<tr>
<td>Nonteaching</td>
<td>72.0 (605)</td>
</tr>
<tr>
<td>Teaching</td>
<td>28.1 (164)</td>
</tr>
<tr>
<td>Primary stroke center%</td>
<td>69.0 (437)</td>
</tr>
</tbody>
</table>

Abbreviations: CAD, coronary artery disease; IQR, interquartile range; MI, myocardial infarction; NIHSS, National Institutes of Health Stroke Scale; TIA, transient ischemic attack.

aArrival outside of regular work hours from 7 AM to 6 PM Monday through Friday.

bCalculated as weight in kilograms divided by height in meters squared.

cIncludes patients who transferred out.

dData for these hospital characteristics are given as percentage of patients (No. of hospitals). Stroke discharge data were missing for 15 hospitals, and geographic region and hospital type were missing for 13 hospitals.

As certified by The Joint Commission.
ranking categories changed based on the 30-day mortality models with and without NIHSS score are shown in Table 3. Compared with the 20%/60%/20% rankings generated by the 2 models, the weighted κ was 0.585 (unweighted κ, 0.530), and the model without NIHSS score differed in category for 206 of 782 hospitals (26.3%) (Table 3).

There was considerable disagreement between the models with and without NIHSS score regarding which hospitals were in the top 5 percentile with respect to the lowest risk-adjusted mortality with weighted κ of only 0.533 (unweighted κ, 0.521). Of the 39 hospitals identified as top performing (top 5 percentile) by the model without NIHSS score, only 23 were identified by the model with NIHSS score, and 16 other hospitals were newly identified (Table 3).

There was even greater disagreement about the bottom-performing hospitals. Of the 40 bottom-performing hospitals according to the claims model without NIHSS score, 19 of these hospitals (47.5%) were no longer identified as being in the worst fifth percentile for mortality after applying the model that adjusted for NIHSS score. For the analysis of hospitals ranked as better than, no different than, and worse than expected for 30-day risk-standardized mortality, 9 of 22 hospitals (40.9%) identified as having better than expected mortality rates in the model without NIHSS score were reclassified to as-expected mortality rates in the model with NIHSS score, and 15 of 26 hospitals (57.7%) classified as worse than expected outcomes were reclassified as to as expected (Table 3). The weighted κ was 0.502 and unweighted κ, 0.494.

To evaluate whether there was selection bias introduced by analyzing the population of acute ischemic stroke patients with NIHSS score documented, the discrimination of the claims model in the overall population (n = 252 379) with or without NIHSS score documented was evaluated. The C statistic (0.772; 95% CI, 0.769-0.774) was similar to that demonstrated for the patients with NIHSS score documented.

**COMMENT**

Using data from the GWTG-Stroke registry combined with administrative

### Table 2. Performance of 30-Day Mortality Risk Models for Acute Ischemic Stroke Without and With NIHSS Score

<table>
<thead>
<tr>
<th></th>
<th>C Statistic (95% CI)</th>
<th>Generalized R²</th>
<th>Predicted Event Rate by Decile of Predicted Risk, %</th>
<th>NRI Index, %</th>
<th>IDI Index, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk model without NIHSS score</td>
<td>0.772 (0.769-0.776)</td>
<td>0.174</td>
<td>Lowest 5.94, Highest 47.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk model with NIHSS score</td>
<td>0.864 (0.861-0.867)</td>
<td>0.335</td>
<td>4.81, 66.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference (95% CI)</td>
<td>0.091 (0.088-0.094)</td>
<td>0.161</td>
<td>93.1 (91.6-94.6), 15.0 (14.6-15.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** IDI, integrated discrimination improvement; NIHSS, National Institutes of Health Stroke Scale; NRI, net reclassification improvement.

*Age, sex, medical history of prior stroke or transient ischemic attack, and 84 condition categories listed in eTable 2 are adjusted for in the modeling.

*Discrimination slope defined as difference of estimated mean probability for events and estimated mean probability for nonevents.

### Table 3. Hospital Ranking Agreement Based on 30-Day Mortality Risk Models Without and With Adjustment for NIHSS Score

<table>
<thead>
<tr>
<th>Rank Based on Model With NIHSS Score</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 20%</td>
<td>156</td>
</tr>
<tr>
<td>Middle 60%</td>
<td>469</td>
</tr>
<tr>
<td>Bottom 20%</td>
<td>157</td>
</tr>
<tr>
<td>Total</td>
<td>782</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank Based on Model Without NIHSS Score</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 5%</td>
<td>39</td>
</tr>
<tr>
<td>Middle 90%</td>
<td>703</td>
</tr>
<tr>
<td>Bottom 5%</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>782</td>
</tr>
</tbody>
</table>

**Abbreviation:** NIHSS, National Institutes of Health Stroke Scale.

*Hospitals with the 95% credible intervals of the estimated random intercepts not covering the null point are considered to have performance that is significantly better or worse than the average hospital. These categories are analogous to the portions of hospitals identified in Hospital Compare as having better than, no different than, or worse than expected 30-day risk-standardized mortality rates.

©2012 American Medical Association. All rights reserved.
outcomes.\textsuperscript{11} Whether or not clinicians contributed to better or worse patient approaches to patient management con-

for patients with less severe illness, re-

prospective risk-adjusting the re-

comparisons should be done only af-

outcomes are to be compared, such

prospective risk-adjusting the re-

for patients with differing case mix and ill-

risks for acute ischemic stroke for Medicare ben-

ficiaries. Furthermore, this study sug-

suggests that inclusion of admission stroke severity may be essential for optimal ranking of hospital with respect to 30-

day mortality.

The increasing use of 30-day mor-

tality rates to assess hospital quality has intensified their importance. This outcome measure has the potential to give both patients and clinicians important feedback concerning a hospital’s qual-

ity of care.\textsuperscript{5} Because hospitals may treat patients with differing case mix and illness severity, adequate risk adjust-

ment is essential.\textsuperscript{11} Outcome measures that do not adequately risk-adjust may systematically favor hospitals that care for patients with less severe illness, regardless of whether these hospitals’ ap-

proaches to patient management contributed to better or worse patient outcomes.\textsuperscript{11} Whether or not clinicians agree with the basic tenets of hospital comparisons, nearly all agree that, if outcomes are to be compared, such comparisons should be done only af-

after appropriately risk-adjusting the re-

sults.\textsuperscript{11,24,25} The primary concern with outcome measures that do not ade-

equately discriminate mortality risk is that hospital rankings based on these models may distort hospital profiling and quality assessment.\textsuperscript{24,25} Thus, a key question confronting clinicians, hos-

pitals, payers, and policy makers is whether current and emerging mea-

sures that assess 30-day mortality are adequate for public reporting and use for rewarding and penalizing hospi-

tals in value-based purchasing.\textsuperscript{7}

It has been previously reported that risk-standardization models for non-

stroke conditions that adjust for demo-

graphics and comorbid conditions based on administrative claims data are sufficient for public reporting, despite not adjusting for indicators of disease severity, laboratory test results, and di-

agnostic studies at time of presenta-

tion.\textsuperscript{5,6} The CMS is now considering an outcome measure of 30-day mortality for acute ischemic stroke.\textsuperscript{26}

In this GWTG-Stroke analysis, we demonstrate that a hospital risk model based on claims data alone without ad-

justment for stroke severity has sub-

stantially worse discrimination com-

pared with a model that adjusts for stroke severity using the NIHSS score. This study also suggests that the ranking of hospitals could be confounded if the risk-adjustment models do not take into account the severity of the acute ischemic stroke, demographics, and other factors present at the time of acute ischemic stroke presentation. For conditions such as heart failure, acute myocardial infarction, and pneumo-

nia, it is believed that claims-only mod-

els for 30-day mortality can ade-

quately discriminate mortality risk at the hospital level for Medicare pa-

tients.\textsuperscript{3,5} However, in contrast to these conditions, these findings suggest that this is not the case for hospital risk mod-

els for acute ischemic stroke.

It logically follows that a measure of stroke severity would be essential for optimal discrimination of mortality risk, as NIHSS score is well documented to be a key risk determinant in acute is-

chemic stroke.\textsuperscript{12,13} Prior patient-level analyses have shown that NIHSS score was the strongest predictive variable for in-hospital and 30-day mortality and substantially improved the performance of a model based on clinical vari-

ables without stroke severity.\textsuperscript{14,15} Our findings are also consistent with, and substantially extend, prior smaller and regionally restricted analyses. An analy-

sis of 2 administrative data prediction models used to assess New York hos-

pitals found that, in the absence of a measure of index stroke severity, the mortality prediction models were non-

congruent and yielded hospital rank-

ings that agreed only slightly more of-

ten than expected by chance.\textsuperscript{24}

As public reporting and value-based purchasing policies increase for out-

come measures, it is important to rec-

ognize the effect that using models with less than ideal discrimination and cali-

bration has on the ranking of hospitals and the lack of correlation among rank-

ing by models that do and do not adjust for critical risk determinants.\textsuperscript{7} To truly identify highest- and lowest-perform-

hospitals for acute ischemic stroke outcomes, models that adjust for stroke severity will likely be needed.

It is also important to carefully con-

sider that rewarding or punishing hos-

pitals on the basis of a risk model that does not account for stroke severity may misalign incentives, as many hospitals identified as performing better than or worse than expected were likely mis-

classified. As a consequence, hospitals may consider turning away patients with more severe strokes or transferr-

ing them to other hospitals after emer-

gency department assessment to avoid being misclassified as having higher risk-standardized mortality. A 30-day mortality model for acute ischemic stroke without adjustment for stroke se-

verity provides lesser discrimination, produces different rankings of hospi-

tal performance, and may be biased in favor of hospitals treating less severe strokes than a model with adjustment for stroke severity.

Although the present analysis sup-

ports collection and adjustment of stroke severity for hospital 30-day mor-
tality, NIHSS score was recorded for 50.7% of hospitalized acute ischemic stroke patients during the study period. The time and expertise needed to perform even a short standardized stroke severity assessment and ensuring these data are accurately abstracted and entered into the Hospital Compare data collection system are important barriers that will need to be overcome. The increasing use of quality measures reporting the portion of acute ischemic stroke patients with NIHSS scores recorded may facilitate this process. The Hospital Alliance and CMS should consider requiring the collection and reporting of the initial NIHSS score for all hospitalized acute ischemic stroke patients prior to implementation of a 30-day mortality risk-standardized model.

Several limitations in our study should be noted. First, the patient population studied is Medicare fee-for-service beneficiaries enrolled in GWTG-Stroke and may not be representative of all patients hospitalized with acute ischemic stroke. However, the differences between patients with and without NIHSS score recorded were small, and other recent analyses suggest that fee-for-service Medicare patients enrolled in GWTG-Stroke are representative of the entire US population of fee-for-service Medicare patients. Second, this study includes only patients in fee-for-service Medicare and thus does not include patients who were enrolled in managed care, uninsured individuals, and patients younger than 65 years of age. Third, variables used for risk adjustment were drawn from claims data and are dependent on their accuracy, as is any model derived from Medicare administrative claims databases. Fourth, these risk models did not adjust for therapies provided, such as tissue plasminogen activator. Fifth, as is the convention for CMS mortality models, the primary outcome was all-cause mortality rather than cause-specific mortality (eg, stroke-specific mortality). We did not assess 30-day rehospitalization, health-related quality of life, functional recovery, patient satisfaction, and other clinical outcomes that may be of interest for hospital outcome measures, and the need for adjustment for stroke severity for these other outcome measures requires further study. Sixth, the hospital ranking methodologies applied did not use shrinkage analysis and may have differed in other ways than those currently applied by CMS. Whether these differences would affect these findings needs additional analysis, but prior studies have suggested selection of risk adjustors had far more influence on ranking profiles than choice of statistical strategies.

CONCLUSIONS

Adding stroke severity assessed with the NIHSS score to a hospital 30-day mortality model based on claims data for Medicare beneficiaries with acute ischemic stroke is associated with substantial improvement in model discrimination and changes in mortality performance ranking for a considerable proportion of hospitals. These findings suggest that it may be critical to collect and include stroke severity for optimal hospital risk adjustment of 30-day mortality for Medicare beneficiaries with acute ischemic stroke.

Author Contributions: Dr Fonarow had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Fonarow, Saver, Reeves, Kleindorfer, Peterson.

Acquisition of data: Fonarow, Saver, Peterson.

Analysis and interpretation of data: Fonarow, Pan, Saver, Smith, Reeves, Broderick, Kleindorfer, Sacco, Olson, Hernandez, Peterson, Schwamm.

Drafting of the manuscript: Fonarow.

Critical revision of the manuscript for important intellectual content: Fonarow, Pan, Saver, Smith, Reeves, Broderick, Kleindorfer, Sacco, Olson, Hernandez, Peterson, Schwamm.

Statistical analysis: Pan, Saver, Peterson.

Obtained funding: Fonarow.

Administrative, technical, or material support: Fonarow, Saver, Kleindorfer, Olson, Hernandez, Schwamm.

Study supervision: Fonarow, Olson, Schwamm.

Conflicts of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Fonarow reported serving as a member of the Get With The Guidelines (GWTG) Steering Committee and receiving research support from the National Institutes of Health and an employee of the University of California, which holds a patent on retriever devices for stroke. Dr Pan is a member of the Duke Clinical Research Institute (DCRI), which serves as the American Heart Association (AHA) GWTG data coordinating center. Dr Saver reported serving as a member of the GWTG Science Oversight Committee and as a consultant regarding trial design and conduct to Covidien, CoAxia, Talascri, Braintree, Syngis, and Ev3 and is an employee of the University of California, which holds a patent on retriever devices for stroke. Dr Smith reported having served on an advisory board to Genentech and is on the data safety and monitoring board for the MR WITNESS trial. Dr Reeves reported receiving salary support from the Michigan Stroke Registry and serving as a member of the AHA GWTG Quality Improvement Subcommittee. Dr Broderick reported receiving funding from the National Institute of Neurological Disorders and Stroke (NINDS) for multiple ongoing trials; receiving study medications from Genentech for 2 ongoing NINDS studies and having received research payment as a consultant to Genentech; and having been reimbursed for travel to meetings by Genentech. Dr Sacco is immediate past president of the AHA. Dr Olson is a member of the DCRI, which serves as the AHA GWTG data coordinating center, and reported serving as a consultant to Bristol Myers Squibb/Sanoﬁ. Dr Hernandez is a member of the DCRI, which serves as the AHA GWTG data coordinating center, and reported being a recipient of an AHA Pharmaceutical Roundtable grant and having received research support from Johnson & Johnson and Amlyn. Dr Peterson reported serving as principal investigator of the Data Analytic Center for AHA GWTG; reported receiving research grants from Johnson & Johnson, Eli Lilly, and Janssen Pharmaceuticals; and reported serving as a consultant to Boehringer Ingelheim, Johnson & Johnson, Medscape, Merck, Novartis, Orpho-McNeil-Janssen, Pfizer, Westat, the Cardiovascular Research Foundation, WebMD, and United Healthcare. Dr Schwamm reported serving as chair of the AHA GWTG Steering Committee and as a consultant to the Massachusetts Department of Public Health. No other disclosures were reported.

Funding/Support: The GWTG-Stroke program is provided by the AHA/American Stroke Association. The GWTG-Stroke program is currently supported in part by a charitable contribution from Janssen Pharmaceutical Companies of Johnson & Johnson. GWTG-Stroke has been funded in the past through support from Boehringer-Ingelheim, Merck, Bristol-Myers Squibb/Sanoﬁ, Pharmaceutical Partnership, and the AHA Pharmaceutical Roundtable.

Role of the Sponsor: The industry sponsors of GWTG-Stroke had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.

Disclaimer: Dr Peterson, Contributing Editor for JAMA, was not involved in the editorial review of the decision to publish this article.


REFERENCES


©2012 American Medical Association. All rights reserved.
Hospital Performance in Acute Ischemic Stroke


