__Supplemental Report to NQF Neurology Steering Committee__

The following document provides clarifications regarding the publication below:

> Fonarow GC, Pan W, Saver JL, Smith EE, Reeves MJ, Broderick JP, Kleindorfer DO, Sacco RL, Olson DM, Hernandez AF, Peterson ED, Schwamm LH. Comparison of 30-day mortality models for profiling hospital performance in acute ischemic stroke with vs without adjustment for stroke severity. JAMA. 2012 Jul 18;308(3):257-64.

Responses to criticisms/concerns raised with the publication.

NIHSS is missing in half the patients and this may impact the findings

The hospital characteristics are very similar for those patients with NIHSS not recorded and those with NIHSS. These two populations also have similar patient characteristics. Further the 30-day mortality rates are similar (14.49% vs. 14.94% for those patients with NIHSS recorded and those without NIHSS, respectively). The hospital NIHSS missing rate and hospital median/mean NIHSS score only have only a mild positive relationship. The reclassification of hospitals was not associated with the NIHSS missing rate. Hospitals' NIHSS missing rate and hospitals' 30-day risk standardized mortality rates had a very negligible positive relationship. There is no valid data which suggests that missing NIHSS in a portion of patients had any meaningful impact on these findings. Had data on NIHSS been available in all patients, it is likely that even greater reclassification of hospitals would have been demonstrated.

Hemorrhage stroke patients were included in the JAMA paper

Only patients with acute ischemic stroke were included in the JAMA paper as clearly noted throughout the manuscript. The demographics, clinical characteristics, in-hospital and 30-day mortality of the JAMA cohort of patients precisely match that of the national Medicare population with acute ischemic stroke.

The JAMA paper included different variables and is not parsimonious

The variables selected closely matched the Yale/CMS model for all variables except for one covariate added after the public comment period. Reducing the variables down to a smaller number and more parsimonious model, made no differences in the model performance for hospital ranking and slightly reduced model discrimination.

There has been no evidence provided that including patients transferred in for the index acute ischemic stroke admission from an outside Emergency Department improves model discrimination to match that obtained with the addition of NIHSS score, reduces hospital misclassification, or is an adequate substitute for adjusting for NIHSS.

In many communities patients with more severe stroke are directly transported by Emergency Medical Services to primary stroke centers or comprehensive stroke centers bypassing closer hospitals. This is not captured or adjusted for in the Yale/CMS model, yet greatly influences the case mix/stroke severity for hospitals.

Having evaluated clinical and administrative data separately may have produced different findings

The impact of adding NIHSS to a parsimonious clinical variable based model for 30-day risk standardized mortality rate was separately reported in a publication in the *Journal of the American Heart Association (J Am Heart Assoc* 2012, 1:42-50). A derived and validated parsimonious clinical model was substantially outperformed by NIHSS score alone (c-statistic 0.71 clinical model vs. NIHSS alone c-statistic 0.82, P<0.0001). Further when comparing the clinical model alone to a combined clinical-NIHSS based model discrimination improved markedly with the addition of NIHSS (c-statistic 0.71 clinical model vs. c-statistic 0.84 combined clinical model-NIHSS, P<0.0001). Whether clinical or administrative based risk models are analyzed and irrespective of which analytic approaches are used, adjusting for stroke severity with NIHSS substantially improves model discrimination and prevents substantial misclassification of hospital performance.

Adding an additional covariate will reduce overall variance at the hospital level and have the effect of pulling some outliers in as seen in the reclassification analysis

No other administrative or clinically derived variable could be identified which had this substantial impact on model discrimination or produced this difference in variance. There was 36-40% of the variation in risk standardized mortality rates (RSMR) on the model based on administrative claims data with adjustment of NIHSS that is unexplained by the model based on administrative claims data without adjustment for NIHSS.

Comparison of intercepts does not capture the full case mix of the hospitals as a risk-standardized rate

The method employed in the JAMA publication ranked each hospital based on the difference between its predicted hospital mortality and its expected hospital mortality. The higher the difference is, the worse this hospital performs. This difference, represented by the random intercept of the hierarchical modeling, follows a Gaussian distribution. If the 95% confidence interval of this difference covers the null (0), then this hospital meets its expected mortality; otherwise if the 95% CI of this difference is above the null, then this hospital performs significantly worse than expectation; or if the 95% CI of this difference is below the null, then this hospital performs significantly better than expectation.

The Yale/CMS models ranks a hospital based on the ratio of its predicted hospital mortality and its expected hospital mortality (times the national average, which is a constant). The higher the ratio is, the worse this hospital performs. The estimated confidence interval of this ratio is obtained by bootstrapping technique. As in the JAMA publication method, the classification is made according to the coverage of the ratio's 95% confidence intervals on the national average for 30-day mortality. These two methods are statistically equivalent.

As many as 58% of the hospitals identified as having better than or worse than expected mortality based on a risk model that does not adjust for stroke severity would be reclassified to as expected mortality if risk adjustment were based on a model that does adjust for stroke severity with the NIHSS.

The JAMA article does not allow evaluation of the degree of differences between the two models

The JAMA publication provides multiple parameters comparing models with and without adjustment for NIHSS score including evaluating the impact on model discrimination, correlation coefficients, median change in hospital rank, net reclassification improvement, integrated discrimination improvement, and 3 different methods for hospital performance ranking.

To provide additional perspective of the degree of differences between the two models, the differences in absolute terms between hospital 30-day risk-standardized mortality rates with and without adjustment for stroke severity are provided below.

- # (percent) hospitals with the two 30-day RSMRs differ by 1 percent or more in absolute terms: 239/782 (30.56%)

- # (percent) hospitals with the two 30-day RSMRs differ by 2 percent or more in absolute terms: 52/782 (6.65%)

A substantial portion of hospitals will have their RSMR performance for 30 day mortality in acute ischemic stroke mischaracterized in absolute terms and in relative performance ranking terms by virtue of not accounting for stroke severity.

The JAMA paper does not present the correlation between the original model results and new results for hospitals

The correlation of the RSMRs based on 30-day mortality models with and without NIHSS is 0.796 ($p<0.0001$, Pearson's correlation) or 0.775 ($p<0.0001$, Spearman's correlation). As expected, there is some correlation between the RSMRs, but this is well below the levels of correlation that have previously been reported for comparing claims based models to clinical models.

As a matter of routine, it is the squared correlations which should be interpreted. The squared correlation describes the proportion of variance in common between the two models. The R^2 is 0.633 or 0.601. This implies that 36-40% of the variation in RSMRs in the model with NIHSS is unexplained by the model without NIHSS.

We also compared hospitals' RSMR based on the model without NIHSS adjustment to the RSMR based on the model with NIHSS adjustment in terms of the sample standard deviation. The standard deviation for RSMR is 0.01767 without NIHSS and 0.01524 with NIHSS. The standard deviation is reduced by 13.8% (variance reduced by 25.6%) providing further evidence on how model fitting improved by adjusting for NIHSS score.

<u>Data from GWTG-Stroke may not be as reliable as that obtained from administrative data</u>

GWTG-Stroke data is subject to systematic efforts and auditing for data quality assurance. All sites and chart abstractors receive ongoing training. The GWTG-Stroke data abstraction tool includes predefined logic features and user alerts to identify potentially invalid data formats or values. Required fields are structured so that valid data must be entered before the data can be saved as a complete record and entered into the database. Edit checks are used to identify inconsistent or out-of-range data and prompt the user to correct or review data entries that are outside a predefined range. Feedback on completeness and quality of the data are provided to participating hospitals on a regular basis. The accuracy and reliability of data entered in GWTG-Stroke have been evaluated with the results published in the peer review literature. Data entered by sites in the GWTG-Stroke database were compared with that abstracted from de-identified medical records by trained auditors. Accuracy for each individual data element and a composite accuracy measure were calculated. The overall composite accuracy rate was 96.1%. The initial NIHSS score was demonstrated to have an accuracy of 93.6%, with an excellent inter-rater reliability of 0.89 (kappa statistic, 95% confidence interval 0.79-0.91). This audit establishes the reliability of GWTG-Stroke registry data in general and for NIHSS specifically.

Any imprecision or lack of uniformity in NIHSS scoring would be expected to diminish, not enhance, the prognostic value of NIHSS.

Xian Y, Fonarow GC, Reeves MJ, Webb LE, Blevins J, Demyanenko VS, Zhao X, Olson DM, Hernandez AF, Peterson ED, Schwamm LH, Smith EE. Data quality in the American Heart Association Get With The Guidelines-Stroke (GWTG-Stroke): results from a national data validation audit. Am Heart J. 2012 Mar;163(3):392-8, 398.e1.