Stroke Severity and Comorbidity Index for Prediction of Mortality after Ischemic Stroke from the Virtual International Stroke Trials Archive–Acute Collaboration

Thanh G. Phan, FRACP, PhD, Benjamin Clissold, FRACP, John Ly, FRACP, Henry Ma, FRACP, Chris Moran, FRACP, Velandai Srikanth, FRACP, PhD on behalf of the VISTA–Acute Collaboration[1]

Background: There is increasing interest in the use of administrative data (incorporating comorbidity index) and stroke severity score to predict ischemic stroke mortality. The aim of this study was to determine the optimal timing for the collection of stroke severity data and the minimum clinical dataset to be included in models of stroke mortality. To address these issues, we chose the Virtual International Stroke Trials Archive (VISTA), which contains National Institutes of Health Stroke Scale (NIHSS) on admission and at 24 hours, as well as outcome at 90 days. *Methods:* VISTA was searched for patients who had baseline and 24-hour NIHSS. Improvement in regression models was performed by the net reclassification improvement (NRI) method. *Results:* The clinical data among 5206 patients were mean age, 69 ± 13; comorbidity index, 3.3 ± .9; median NIHSS at baseline, 12 (interquartile range [IQR] 8-17); NIHSS at 24 hours, 9 (IQR 8-15); and death at 90 days in 15%. The baseline model consists of age, gender, and comorbidity index. Adding the baseline NIHSS to model 1 improved the NRI by 0.671 (95% confidence interval [CI] 0.595-0.747) [or 67.1% correct reclassification between model 1 and model 2]. Adding the 24 hour NIHSS term to model 1 (model 3) improved the NRI by 0.929 (95% CI 0.857-1.000) for model 3 versus model 1. Adding the variable thrombolysis to model 3 (model 4) improve NRI by 0.1 (95% CI 0.023-0.178) [model 4 versus model 3]. *Conclusion:* The optimal model for the prediction of mortality was achieved by adding the 24-hour NIHSS and thrombolysis to the baseline model. **Key Words:** Ischemic stroke—mortality—Charlson Comorbidity Index—prognosis.

From the Stroke Unit, Monash Health and Stroke and Aging Research Group, Monash University, Melbourne, Australia.
Received June 25, 2015; revision received October 23, 2015; accepted December 9, 2015.

Disclosure: Professor Phan reports receiving honoraria as speaker for Genzyme, Boehringer Ingelheim, and Bayer. He is on the advisory board for Genzyme on Fabry disease.

Author contributions: Study concept and design: T.G.P., B.C., H.M., J.L., V.S.; acquisition of data: T.G.P., VISTA–Acute collaborators; analysis and interpretation of data: T.G.P., B.C., and C.M.; drafting of the manuscript: T.G.P., B.C., H.M., J.L., V.S.

Address correspondence to Thanh G. Phan, FRACP, Stroke and Aging Research Group, Department of Medicine, Monash University, Vic. 3168, Australia. E-mail: Thanh.Phan@monash.edu.

[1] VISTA–Acute Steering Committee: K.R. Lees (Chair), A. Alexandrov, P.M. Bath, E. Bluhmki, N. Bornstein, L. Claesson, S.M Davis, G. Donnan, H. C. Diener, M. Fisher, M. Ginsberg, B. Gregson, J. Grotta, W. Hacke, M.G. Hennerici, M. Hommel, M. Kaste, P. Lyden, J. Marler, K. Muir, R. Sacco, A. Shuaib, P. Teal, N.G. Wahlgren, S. Warach, and C. Weimar.

1052-3057/$ - see front matter

http://dx.doi.org/10.1016/j.jstrokecerebrovasdis.2015.12.016

Introduction

Headlines about hospital performances sometimes dominate the front page of newspapers, with particular attention around the issue of hospital mortality. Globally, administrators and hospital organizations are justifiably concerned and are looking at different approaches to measure hospital performance in the hope that better measurement would reflect improvement in care. Mortality is often the preferred outcome because of the ease of measurement of this metric. Due to concerns with the use of raw mortality data, the concept of standardized mortality ratio has been used.[1] The standardized mortality ratio is the observed mortality divided by the expected mortality calculated from data on comorbidity.[1,2] This approach is attractive because it uses routinely collected information available in hospital administrative data.

Investigators from Get With the Guidelines-Stroke in North America[3] have shown that stroke severity measured using the National Institutes of Health Stroke Scale (NIHSS) is a predictor of mortality from stroke in addition to comorbidity index. In that study, the timing of the stroke severity collection was not specified other than that it was based on the first recorded NIHSS score, judged to be "as close to the admission as possible."[3] In the absence of a strict definition of the time for NIHSS measurement in this paper, this could either be from 1 hour to several days later. It is unknown whether the timing of measurement of stroke severity, or indeed other clinical variables, influences the accuracy of prediction of mortality. Further, it is also not known if other clinical data (such as the use of recombinant tissue plasminogen activator [rTPA]) would be useful in this regard. The collection of such data is logistically difficult in real-world clinical practice outside of randomized trials. To address these issues, we chose the Virtual International Stroke Trials Archive (VISTA) dataset, which contains data on NIHSS score on admission and at 24 hours, as well as outcome at 90 days.[4]

The aim of the present study was to determine the optimal timing for collection of stroke severity data and the minimum clinical dataset to be included in models of mortality. Such analyses are important in setting the boundary for the collection of the minimum dataset.

Methods

The VISTA archive contains data from completed randomized clinical trials and registry of stroke patients (http://www.vista.gla.ac.uk/).[4,5] The VISTA archive contains 8 subcommittees: Acute, Prevention, Rehabilitation, Intracerebral Hemorrhage, Imaging, Plus, Endovascular, and Cognition. Registry data and observational studies are held in VISTA–Plus. The data from the current study are from VISTA–Acute, which contains data from completed randomized clinical trials. The data are released in a deidentified manner so that the trials and treatment allocations are not known. This approach was used to facilitate "novel exploratory analysis of data" via access to a large volume of high-quality clinical trial data.[5] It was hoped that such pooling of data can be used to plan for future clinical trials. For the present study, we first submitted a proposal to the VISTA steering committee, which took appropriate steps to ensure that the project was not a duplication of existing or previously published VISTA projects.[6,7]

We searched the VISTA archive for the following: imaging data NIHSS score on admission and at 24 hours; physiological variables (systolic blood pressure and blood glucose level); demographic data (age, gender); stroke risk factors and comorbidity (including but not limited to hypertension, diabetes, atrial fibrillation, degree of liver impairment, and degree of renal impairment); thrombolysis treatment with rTPA—Alberta Stroke Program Early CT Score (ASPECTS) score on CT scan; clinical data; and outcome of death within 90 days of stroke.

Charlson Comorbidity Coding in VISTA

The Charlson Comorbidity Index is derived from administrative coding of hospital data.[8] It can be conceived as a weighted index of comorbidity conditions. For example, stroke is assigned a weighting of 1, but hemiplegia is assigned a weighting of 2. In the present study, the variable diabetes was given a comorbidity coding of 2 because stroke represents "diabetes with end organ damage." To code motor deficit for the Charlson Comorbidity Index, we used an NIHSS score equal to or greater than 6 (the minimum NIHSS score in these trials was 6). Weighting of complex illness may vary; for example, severe liver disease is assigned a weighting of 3, whereas acquired immune deficiency syndrome and metastatic cancer have a weighting of 6. Patients with these conditions are generally not included in clinical trials of stroke and are under-represented in VISTA. In this analysis, the comorbid conditions were summed together.

Statistical Analysis

We performed logistic regression analyses in several stages to derive the optimal model that requires the minimum number of inputs.

Model 1 = age + male gender + Charlson Comorbidity Index

Model 2 = model 1 variables + admission NIHSS score

Model 3 = model 2 variables + 24-hour NIHSS score

Model 4 = model 3 variables + rTPA

Model 5 = age + 24-hour NIHSS score

Model 6 = model 4 + other physiological variables, risk factors, and imaging data, including systolic blood pressure, serum glucose level, hypertension, atrial fibrillation, and ASPECTS score.

Assessing Model Discrimination

The areas under the receiver operating characteristic curve were used to assess how well the models discriminate between those who died and those who survived (at days 7, 30, and 90).

Model Calibration

We performed this by using the Hosmer–Lemeshow goodness-of-fit test,[9] and the Nagelkerke generalized R^2.[10] A model is well calibrated when the Hosmer–Lemeshow goodness-of-fit test shows no difference between observed and expected outcome or *P* value approaching 1. A high generalized R^2 value suggests a well-calibrated regression model.

Measuring Improvement in Regression Models

Due to the low sensitivity of areas under the curve (AUCs) for detecting differences in discrimination between models, the net reclassification improvement (NRI) and integrated discrimination improvement (IDI) have been proposed as more sensitive metrics of improvement in model discrimination.[11] The NRI can be considered as a percentage reclassification for the risk categories and the IDI is the mean difference in predicted probabilities between 2 models (constructed from cases with disease and without disease). The NRI and IDI scores are expressed here as fractions and can be converted to percentage by multiplying 100. The continuous NRI and IDI were performed using PredictABEL (R Statistical Foundation, https://www.r-project.org/).[12]

Results

The demographic data (Table 1) among 5206 patients were as follows: mean age 68.8 ± 12.5 with 55% males, 54.5% ever-smokers, 73.6% with hypertension, 23% with diabetes. The comorbidity score (based on a baseline NIHSS score ≥6) was 4.2 ± .84. The proportion of subjects receiving rTPA was 41.0%. The mean NIHSS score was 12.9 ± 5.3 at baseline and 10.5 ± 7.2 at 24 hours (the proportion of patients with an NIHSS score of 20 or greater was 16.1%). Death within 7 days occurred in 4.6% of subjects and increased to 10.6% at 30 days and 15.1% at 90 days. In this analysis, only 1807 of the 5206 subjects had an ASPECTS score recorded with a mean ASPECTS score of 9.8 ± .8 (among these patients with ASPECTS score available, the onset to treatment time was 3.5 ± .8 hours).

Univariable analyses showed statistically significant relationships between mortality and NIHSS score at baseline and at 24 hours, age, systolic blood pressure, atrial fibrillation, and the use of rTPA. The multivariable models for mortality and their associated AUC, Hosmer–Lemeshow goodness-of-fit, and generalized R^2 are displayed in Table 2. The baseline model (model 1) incorporating age and comorbidity index had an AUC of .695 (95% confidence interval [CI] .674-.716). There is improvement in the AUC with successive change to the model (see Table 2). The optimal model at 90 day was model 4 (age, male gender, comorbidity index, NIHSS score at 24 hours and rTPA). The AUC for this model was .830 (95% CI .814-.846).

Table 1. *Patient demographics*

	N	Mean
Age	5206	68.8 ± 12.5
Gender	5206	.55 ± .5
rTPA	5206	.41 ± .49
AF	5206	.26 ± .04
ASPECTS	1807	9.8 ± .8
Baseline NIHSS score	5206	12 (IQR 8-17)
NIHSS score, 24 h	5206	9 (IQR 5-15)
mRS score (>2)	5206	.45 ± .49
Comorbidity	5206	4.20 ± .84
Mortality, day 7	5206	.05 ± .21
Mortality, day 30	5206	.11 ± .31
Mortality, day 90	5206	.15 ± .36

Abbreviations: AF, atrial fibrillation; ASPECTS, Alberta Stroke Program Early CT Score; IQR, interquartile range; mRS, modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale; rTPA, tissue plasminogen activator.

Adding baseline NIHSS score to model 1 improved the NRI score to .671 (95% CI .595-747) (or 67.1% correct reclassification as 1 move from model 1 to model 2). Adding the 24-hour NIHSS term to model 1 (model 3) improved the NRI score to .929 (95% CI .857-1.000) (or 92.9% correct reclassification as 1 move from model 1 to model 3). The measurement of improvement between models 3 and 4 was statistically different: the NRI score was .100 (95% CI .023-.178) and the IDI score was .003 (95% CI .00008-.005) (Table 2). In terms of calibration, there was a similar improvement in the Hosmer–Lemeshow goodness-of-fit test and generalized R^2 values (see Table 2). The plots of these receiver operating characteristic curves are displayed in Figure 1.

Mortality Time Points

The statistically significant covariates for models 1-6 were similar for day 30 and day 90, the exception being that the covariate rTPA was not statistically significant for mortality at day 30 (Table 3). The AUC for model 1 was .684 (95% CI .66-.709). The optimal model at day 30 was model 3 (age, male gender, comorbidity index, and NIHSS score at 24 hours). The AUC was .847 (95% CI .830-.864). The measurement of improvement between models 2 and 3 was statistically different: the NRI score was .802 (95% CI .718-.887) and the IDI score was .095 (95% CI .079-.110) (Table 3).

The covariates for early mortality at day 7 were different from those above (Table 4). The AUC for model 1 was .620 (95% CI .581-.659). The optimal model at day 7

Table 2. *Models of mortality at 90 days*

				Discrimination		Comparison of models			
	Predictors	OR (95% CI)	*P* value	AUC (95% CI)	HL (*P* value)	Generalized R^2	Comparison of models	Continuous NRI (95% CI)	IDI (95% CI)
Model 1	Comorbidity	1.357 (1.243-1.481)	<.001	.695 (.674-.716)	.09	.106			
	Age	1.047 (1.038-1.057)	<.001						
	Gender	1.236 (1.044-1.463)	.01						
Model 2	Comorbidity	1.341 (1.223-1.471)	<.001	.772 (.754-.791)	.5	.213	Model 2–Model 1	.671 (.595-.747)	.079 (.069-.089)
	Age	1.040 (1.031-1.050)	<.001						
	Male Gender	1.370 (1.148-1.636)	<.001						
	Baseline NIHSS score	1.151 (1.132-1.1690	<.001						
Model 3	Comorbidity	1.328 (1.204-1.465)	.01	.828 (.812-.844)	.8	.316	Model 3–Model 1	.929 (.8573-1.000)	.167 (.152-.183)
	Age	1.040 (1.029-1.051)	<.001				Model 3–Model 2	.767 (.694-.839)	.088 (.076-.100)
	Male gender	1.359 (1.128-1.639)	<.001						
	NIHSS score, 24 h	1.161 (1.146-1.176)	<.001						
Model 4	Comorbidity	1.3178 (1.194-1.454)	<.001	.830 (.814-.846)	.9	.319	Model 4–Model 1	.919 (.8477-.991)	.170 (.154-.186)
	Age	1.040 (1.029-1.051)	<.001				Model 4–Model 2	.716 (.643-.789)	.091 (.079-.103)
	Male gender	1.367 (1.134-1.649)	<.001				Model 4–Model 3	.100 (.023-.178)	.003 (.0008-.005)
	NIHSS score, 24 h	1.163 (1.148-1.178)	<.001						
	rTPA	.72 (.596-.874)	<.001						
Model 5	NIHSS score, 24 h	1.161 (1.146-1.176)	<.001	.824 (.808-.840)	.9	.306	Model 5–Model 2	.572 (.496-.648)	.079 (.067-.092)
	Age	1.054 (1.044-1.063)	<.001				Model 5–Model 3	−.132 (−.268-.004)	−.004* (−.007-.00002)
	Male gender	1.319 (1.095-1.588)	.004				Model 5–Model 4	−.193 (−.329-−.057)	−.005 (−.009-−.0007)
Model 6	Comorbidity	1.279 (1.155-1.416)	<.001	.831 (.814-.847)	.8	.321	Model 6–Model 1	.919 (.848-.991)	.172 (.156-.188)
	Age	1.042 (1.032-1.053)	<.001				Model 6–Model 2	.709 (.635-.782)	.093 (.080-.105)
	Male gender	1.367 (1.133-1.649)	.001				Model 6–Model 3	.134 (.056-.211)	.004 (.002-.007)
	NIHSS score, 24 h	1.161 (1.146-1.176)	<.001				Model 6–Model 4	.136 (.059-.214)	.002* (−.0001-.003)
	rTPA	.729 (.602-.883)	<.001				Model 6–Model 5	.326 (.247-.405)	.013 (.008-.018)
	BSL	1.035 (1.005-1.065)	.02						

Abbreviations: AUC, area under the curve; BSL, blood sugar level; CI, confidence interval; HL, Hosmer–Lemeshow goodness-of-fit test; IDI, integrated discrimination improvement; NIHSS, National Institutes of Health Stroke Scale; NRI, net reclassification improvement; OR, odds ratio.

*Not significant.

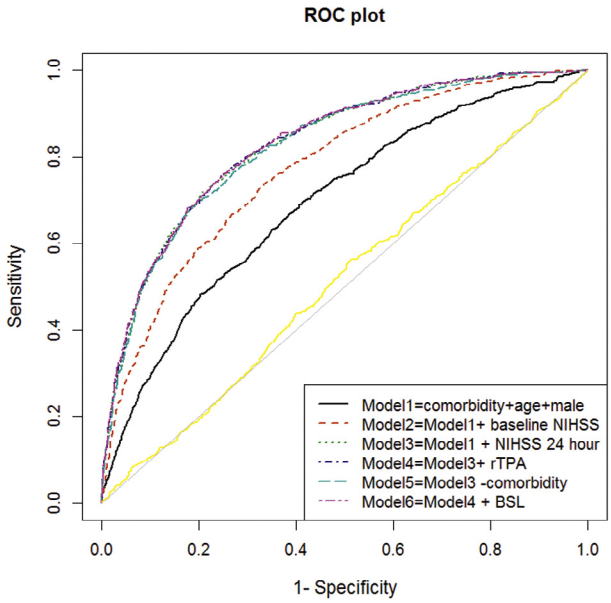

Figure 1. *The different regression models and their associated ROC curves are on display. Abbreviations: BSL, blood sugar level; NIHSS, National Institutes of Health Stroke Scale; ROC, receiver operating characteristic; rTPA, recombinant tissue plasminogen activator.*

was model 3 (age, male gender, comorbidity index, and NIHSS score at 24 hours). The AUC was .863 (95% CI .841-.885). The measurement of improvement between models 2 and 3 was statistically different: the NRI score was .874 (95% CI .749-.999) and the IDI score was .093 (95% CI .069-.116) (Table 4).

Discussion

In the present study, we demonstrate that the use of a stroke severity score measured at 24 hours (rather than at admission) vastly improves the prediction of ischemic stroke mortality at 90 days over and above the predictive ability of age, gender, and comorbidity index. By examining the contribution of other clinical variables, our study supports the concept of a limit on the minimum set of clinical variables that need to be collected for valid prediction of stroke mortality.

The findings on the importance of a stroke severity score such as NIHSS score in the prediction of mortality are consistent with those observed by North American investigators.[3,13,14] The novel finding from our study is that 24-hour stroke severity is superior to admission severity in prediction and may be the best time point for obtaining the NIHSS score for accurately predicting mortality. In clinical practice, this time point corresponds to patients who may often be suspected to have a mild stroke on admission declaring their true clinical outcome at 24-48 hours. This result resonates with the dynamic concept of the ischemic penumbra after stroke and, in some cases, the fate of the ischemic penumbra may only be finalized at 24-48 hours.[15,16]

Related to the above issues is the effect of timing for measuring mortality on covariates such as stroke severity and comorbidity index. The American Heart Association/American Stroke Association had proposed the use of 30-day mortality as the optimal time point for measuring mortality, which coincided with the time point proposed by the Centers for Medicare & Medicaid Services.[17] This shorter time point compared to clinical trial is practical and based on the desire to avoid loss to follow-up.[17] Our findings (Table 3) suggest that the set of predictors used for mortality at day 90 is also valid for day 30.

A strength of our study was the use of a variety of complimentary metrics of discrimination and calibration techniques to evaluate the impact of adding new variables to the regression models. The effect of this can be seen in the comparison between models 3 and 6, where the AUC appeared similar but the NRI and IDI scores were different. This strategy was premised by the low sensitivity of AUC and the high sensitivity of NRI and IDI for detecting changes between models.[18] This strategy showed that the use of rTPA is not discarded from the model if the choice is based on continuous NRI and IDI. The addition of serum glucose level in model 6 resulted in higher NRI score compared to model 4 but not for the IDI score. We can propose that the minimum dataset would include comorbidity, age, male gender, stroke severity at 24 hours, and rTPA.

There are limitations to our approach. The sample we used was obtained from clinical trials and does not fully reflect the breadth of patients usually seen in hospitals. Those excluded from clinical trials may have milder or more severe strokes. The estimation of the comorbidity index in this analysis is also not exactly the same as that derived from usual hospital administrative datasets. Patients in stroke clinical trials do not have significant complex comorbidity such as severe liver failure, terminal malignancy dementia, or human immunodeficiency virus.[8] Further, patients involved in acute stroke trials are more likely to come from tertiary hospitals rather than smaller community hospitals. The VISTA archive contains data on patients included in randomized clinical trials but not characteristics (such as NIHSS score) of patients excluded from these trials.[5] These limitations are likely to underestimate the true impact of stroke severity and other comorbidities on the prediction of mortality. Despite these limitations, the VISTA dataset has the advantage of having a carefully phenotyped comorbidity and other clinical variables (NIHSS score on admission and at 24 hours), whereas hospital administrative datasets may be plagued by inaccuracy in coding of comorbidity and lack other clinical variables that were available for this analysis. The NIHSS score used here was collected prospectively as part of randomized clinical trials. The data are reliable as the vast majority of users of the NIHSS are certified from a single certification window.[19] The video-based

Table 3. *Models of mortality at day 30*

				Discrimination		Comparison of models			
	Predictors	OR (95% CI)	*P* value	AUC (95% CI)	HL (*P* value)	Generalized R^2	Comparison of models	Continuous NRI (95% CI)	IDI (95% CI)
Model 1	Comorbidity	1.320 (1.194-1.460)	<.001	.684 (.66-.709)	.0284	.081			
	Age	1.0446 (1.034-1.056)	<.001						
	Gender	1.207 (.995-1.464)*	.06						
Model 2	Comorbidity	1.292 (1.163-1.436)	<.001	.783 (.763-.803)	.368	.199	Model 2–Model 1	.728 (.642-.815)	.077 (.0661-.088)
	Age	1.036 (1.025-1.047)	<.001						
	Male gender	1.349 (1.101-1.654)	.04						
	Baseline NIHSS score	1.166 (1.144-1.188)	<.001						
Model 3	Comorbidity	1.265 (1.131-1.416)	<.001	.847 (.83-.864)	.216	.313	Model 3–Model 1	1.020 (.940-1.010)	.171 (.153-.190)
	Age	1.036 (1.024-1.048)	<.001				Model 3–Model 2	.802 (.718-.887)	.095 (.079-.110)
	Male gender	1.298 (1.047-1.610)	.02						
	NIHSS score, 24 h	1.174 (1.157-1.191)	<.001						
Model 4	Comorbidity	1.260 (1.125-1.410)	<.001	.848 (.831-.864)	.154	.314	Model 4–Model 1	1.022 (.942-1.101)	.172 (.154-.191)
	Age	1.036 (1.024-1.048)	<.001				Model 4–Model 2	.755 (.670-.840)	.095 (.080-.111)
	Male gender	1.30131.04921.6140	.02				Model 4–Model 3	.041 (−.05-.132)	.0008 (−.0007-.002)
	NIHSS score, 24 h	1.175 (1.158-1.193)	<.001						
	rTPA	.824 (.663-1.024)	<.001						
Model 5	NIHSS score, 24 h	1.175 (1.158-1.192)	<.001	.843 (.826-.861)	.292	.307	Model 5–Model 1	.957 (.8749-1.039)	.166 (.147-.185)
	Age	1.047 (1.036-1.058)	<.001				Model 5–Model 2	.629 (.540-.718)	.089 (.073-.105)
	Male gender	1.266 (1.022-1.568)	.03				Model 5–Model 3	−.146 (−.239-−.054)	−.005 (−.009-−.002)
							Model 5–Model 4	−.244 (−.336-−.151)	−.006 (−.010-−.002)

Abbreviations: AUC, area under the curve; CI, confidence interval; HL, Hosmer–Lemeshow goodness-of-fit test; IDI, integrated discrimination improvement; NIHSS, National Institutes of Health Scale; NRI, net reclassification improvement; rTPA, recombinant tissue plasminogen activator.

Model 6 is the same as model 4 and is not shown.

*Not significant.

Table 4. *Models of mortality at day 7*

				Discrimination		Comparison of models			
	Predictors	OR (95% CI)	*P* value	AUC (95% CI)	HL (*P* value)	Generalized R^2	Comparison of models	Continuous NRI (95% CI)	IDI (95% CI)
Model 1	Comorbidity	1.333 (1.156-1.537)	<.001	.620 (.581-.659)	.388	.0276			
	Age	1.017 (1.003-1.031)	<.001						
	Gender	1.012 (.767-1.334)*	.9						
Model 2	Comorbidity	1.287 (1.110-1.492)	<.001	.78 (.753-.807)	.171	.1374	Model 2–Model 1	.774 (.649-.898)	.044 (.035-.053)
	Age	1.007 (.993-1.021)*	.4						
	Male gender	1.118 (.841-1.487)*	.4						
	Baseline NIHSS score	1.176 (1.146-1.207)	<.001						
Model 3	Comorbidity	1.234 (1.055-1.442)	<.001	.863 (.841-.885)	.119	.2678	Model 3–Model 1	1.118 (1.007-1.228)	.136 (.112-.161)
	Age	1.004 (.989-1.018)*	.6				Model 3–Model 2	.874 (.749-.999)	.093 (.069-.116)
	Male gender	.987 (.730-1.334)*	.9						
	NIHSS score, 24 h	1.185 (1.162-1.207)	<.001						
Model 4	Comorbidity	1.228 (1.051-1.435)	<.001	.863 (.841-.885)	.188	.269	Model 4–Model 1	1.111 (1.000-1.222)	.138 (.113-.163)
	Age	1.003 (.989-1.019)*	.7				Model 4–Model 2	.812 (.684-.939)	.093 (.071-.118)
	Male gender	.989 (.732-1.336)*	.9				Model 4–Model 3	.027 (−.106-.160)	.001 (.0005-.003)
	NIHSS score, 24 h	1.186 (1.164-1.209)	<.001						
	rTPA	.805 (.593-1.093)*	.2						
Model 5	NIHSS score, 24 h	1.186 (1.164-1.209)	<.001	.860 (.838-.882)	.067	.2635	Model 5–Model 1	1.110 (.999-1.220)	.133 (.108-.158)
	Age	1.014 (1.001-1.027)	.04				Model 5–Model 2	.705 (.575-.835)	.089 (.065-.113)
	Male gender	.963* (.713-1.299)	.8				Model 5–Model 3	−.132 (−.2868-.004)*	0-.004* (−.0074-.0004)
							Model 5–Model 4	−.193 (−.329-−.057)	−.005 (−.009-.00007)

Abbreviations: AUC, area under the curve; CI, confidence interval; HL, Hosmer–Lemeshow goodness-of-fit test; IDI, integrated discrimination improvement; NIHSS, National Institutes of Health Scale; NRI, net reclassification improvement; rTPA, recombinant tissue plasminogen activator.

Model 6 is the same as model 4 and is not shown.

*Not significant.

training through this window has been shown to be reliable across multiple venues.[20] In the present study, rTPA is a predictor of mortality at 90 days but not at 7 or 30 days. We have not drawn a strong conclusion on rTPA having an effect on mortality from this finding as the current study was not a randomized control trial. We are reassured that investigators similarly described a recent meta-analysis of thrombolytic trial reduction in mortality at 90 days.[21] In that meta-analysis, death within 7 days was higher in the rTPA arm and was attributed to hemorrhage.[21]

Conclusion

We have determined the optimal time for measuring NIHSS stroke severity and the minimum clinical dataset for reliably predicting stroke mortality. Further, these predictors change depending on the time point for measuring mortality. Using stroke severity in addition to time-appropriate covariates such as age, gender, and comorbidity will enable more valid comparisons of hospital performance.

Acknowledgments: We thank the VISTA investigators and the data custodian for their help with this study.

References

1. Jarman B, Gault S, Alves B, et al. Explaining differences in English hospital death rates using routinely collected data. BMJ 1999;318:1515-1520.
2. Centers for Medicare & Medicaid Services (CMS), HHS. Medicaid program; hospital inpatient value based purchasing program: final rule. Fed Regist 2011;76:26490-26547.
3. Fonarow GC, Pan W, Saver JL, et al. Comparison of 30-day mortality models for profiling hospital performance in acute ischemic stroke with vs without adjustment for stroke severity. JAMA 2013;308:257-264.
4. Ali M, Bath P, Brady M, et al. Development, expansion, and use of a stroke clinical trials resource for novel exploratory analyses. Int J Stroke 2012;7:133-138.
5. Ali M, Bath PM, Curram J, et al. The Virtual International Stroke Trials Archive. Stroke 2007;38:1905-1910.
6. Weimar C, Ziegler A, Sacco RL, et al. Predicting recovery after intracerebral hemorrhage—an external validation in patients from controlled clinical trials. J Neurol 2009;256:464-469.
7. Dowlatshahi D, Demchuk AM, Flaherty ML, et al. Defining hematoma expansion in intracerebral hemorrhage: relationship with patient outcomes. Neurology 2011;76:1238-1244.
8. Charlson ME, Sax FL, MacKenzie CR, et al. Morbidity during hospitalization: can we predict it? J Chronic Dis 1987;40:705-712.
9. Hosmer DW, Hosmer T, Le Cessie S, et al. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med 1997;16:965-980.
10. Nagelkerke NJ. A note on a general definition of the coefficients of determination. Biometrika 1991;78.
11. Pencina MJ, D'Agostino RB Sr, D'Agostino Jr RB, et al. Evaluating the added predictive ability of a new marker: from area under the roc curve to reclassification and beyond. Stat Med 2008;27:157-172, discussion 207-112.
12. Kundu S, Aulchenko YS, van Duijn CM, et al. PredictABEL: an R package for the assessment of risk prediction models. Eur J Epidemiol 2011;26:261-264.
13. Fonarow GC, Saver JL, Smith EE, et al. Relationship of National Institutes of Health Stroke Scale to 30-day mortality in medicare beneficiaries with acute ischemic stroke. J Am Heart Assoc 2013;1:42-50.
14. Saposnik G, Hill MD, O'Donnell M, et al. Variables associated with 7-day, 30-day, and 1-year fatality after ischemic stroke. Stroke 2008;39:2318-2324.
15. Markus R, Reutens DC, Kazui S, et al. Hypoxic tissue in ischaemic stroke: persistence and clinical consequences of spontaneous survival. Brain 2004;127:1427-1436.
16. Ma H, Wright P, Allport L, et al. Salvage of the PWI/DWI mismatch up to 48 h from stroke onset leads to favorable clinical outcome. Int J Stroke 2014;10:565-570.
17. Katzan IL, Spertus J, Bettger JP, et al. Risk adjustment of ischemic stroke outcomes for comparing hospital performance: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2014;45:918-944.
18. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. Ann Intern Med 2009;150:795-802.
19. Lyden P, Brott T, Tilley B, et al. Improved reliability of the nih stroke scale using video training. NINDS TPA Stroke Study Group. Stroke 1994;25:2220-2226.
20. Lyden P, Raman R, Liu L, et al. National Institutes of Health Stroke Scale certification is reliable across multiple venues. Stroke 2009;40:2507-2511.
21. Wardlaw JM, Murray V, Berge E, et al. Recombinant tissue plasminogen activator for acute ischaemic stroke: an updated systematic review and meta-analysis. Lancet 2012;379:2364-2372.