Accuracy of Neurovascular Fellows’ Prognostication of Outcome After Subarachnoid Hemorrhage

Babak B. Navi, MD; Hooman Kamel, MD; Charles E. McCulloch, PhD; Kazuma Nakagawa, MD; Bharath Naravetla, MD; Asma M. Moheet, MD; Christine Wong, MD; S. Claiborne Johnston, MD, PhD; J. Claude Hemphill III, MD, MAS; Wade S. Smith, MD, PhD

Background and Purpose—The purpose of this study was to determine the accuracy and optimal timing of physician prognostication in patients with subarachnoid hemorrhage, a prototypical neurological disease characterized by variable outcomes and frequent disability.

Methods—From October 2009 to April 2010, treating neurologists at a tertiary care academic medical center made daily predictions of the modified Rankin Scale at 6 months for consecutive patients with subarachnoid hemorrhage. Actual functional outcomes at 6 months were determined by phone interview and dichotomized into good (modified Rankin Scale 0–2) and poor (modified Rankin Scale 3–6) outcomes. Descriptive statistics were used to assess the accuracy of prognostications. Multiple logistic regression and generalized estimating equations were used to assess changes in prognostication accuracy over time and the relationship between prognostication accuracy and clinical factors.

Results—Physicians made 648 prognostications for 66 patients. Overall accuracy ranged from 78% to 88%. Among patients predicted to have a good outcome, 81% (95% CI, 71%–92%) actually had a good outcome, whereas 88% (95% CI, 77%–99%) of patients predicted to do poorly had poor outcomes. No significant trends were seen in prognostication accuracy over time during the hospital course ($P=0.72$). Increasing age, infection, mechanical ventilation, hydrocephalus, and seizures all significantly worsened physician accuracy.

Conclusions—Neurologists were generally but not perfectly accurate in their prognostications of functional outcomes. The accuracy of prognoses did not correlate with the hospital day on which they were made but was affected by clinical factors that can cloud the neurological examination. (Stroke. 2012;43:00-00.)

Key Words: outcomes ■ prognosis ■ subarachnoid hemorrhage

Prognostication is an essential role of physicians. Patients and their families require prognostication to emotionally, financially, and socially prepare for death or disability.1–3 The predictions of physicians also allow for patient autonomy, shared decision-making, and assessments of the goals of care. Most studies on physician prognostication have concentrated on mortality predictions in patients admitted to intensive care units or those who are terminally ill with cancer.4–9 Few studies have assessed the accuracy of physicians in predicting outcomes in patients with neurological disease, and even fewer have analyzed the accuracy of neurologists and neurosurgeons in predicting disability.10,11

Crucial family decisions often rely on the prognosis for an acceptable level of neurological functioning rather than on survival, and neurologists are commonly asked to prognosticate about such outcomes, so evaluating their accuracy is particularly important.1,2,12 Subarachnoid hemorrhage (SAH) is an excellent model for studying neurological prognostication because poor outcomes are common, but the course is highly variable with many factors determining the final outcome.13,14 Additionally, prognosticating outcomes after SAH may be more difficult than in other neurological emergencies, because the care of patients with SAH is hindered by overly simplified prediction scores, potentially catastrophic secondary complications, and wide variability in practice patterns.13–16

To address this uncertainty regarding prognostic accuracy after neurological injury, we assessed physician performance in predicting the long-term functional outcomes of patients with SAH. We also examined whether baseline patient characteristics, secondary complications, and discordance in goals of care between treating clinicians and patients or their surrogates affected the accuracy of physician prognostication.
Methods

Study Setting and Design

We performed a prospective cohort study designed to meet the requirements of a quality assurance project with no alteration in patient care and no data collected from patients primarily for research purposes. The University of California, San Francisco neurovascular group undertook the study from October 1, 2009, through March 31, 2010, with the goal of assessing the prognostic accuracy of its physicians. Because this was a quality assurance project, a convenience sample of 6 months was chosen. A mandatory questionnaire was completed on each hospital day for all patients with nontraumatic SAH cared for by the neurovascular service. To ensure full compliance, the questionnaire was integrated into the departmental billing database and completed in conjunction with daily documentation of medical services rendered. The questionnaire was completed by neurovascular fellows, who serve as team leaders on rounds and function as the primary liaisons with patients and their families for all discussions. Four first-year fellows (3 men, 1 woman) and 2 second-year fellows (1 man, 1 woman) participated in the project, although only the fellow on service at any given time would make prognostications (only 1 prognostication each day for each patient). Their mean age was 32 years (range, 30–38 years), and all were neurologists trained in the United States. We report here the results of our analysis of data from this quality assurance project, focusing on prognostications made during the first 14 days of hospitalization. Our institution’s Committee on Human Research approved this analysis. The requirement for informed consent was waived because of minimal risk to patients.

Measurements

A questionnaire documented the physicians’ expectation of the patient’s modified Rankin Scale (mRS) score at 6 months determined prospectively on each day of hospitalization. The mRS is a validated and widely used neurological disability scale that ranges from 0 to 6, with 0 indicating no symptoms and 6 representing death.17 All physicians making predictions had completed an instructional course and were certified in calculating the mRS score before the start of the project. In addition, an electronic link explaining the specific distinctions between mRS scores was placed on the questionnaire.

Additionally, the neurovascular fellows documented the daily goals of care of both the neurovascular team and the patient. In cases in which patients lacked capacity or were unable to communicate, the goals of care of the patient’s healthcare proxy were recorded instead. Options for goals of care for healthcare providers consisted of the following choices: (1) standard care; (2) limited care; or (3) comfort care. Options for goals of care for patients or surrogates consisted of: (1) standard care; (2) limited care; (3) comfort care; (4) unknown because no surrogate was available; or (5) unknown because discussion had not yet occurred.

The following clinical variables were extracted from the electronic medical record of each patient: age, gender, race, medical history (self-reported or documented hypertension, diabetes mellitus, hyperlipidemia, coronary artery disease, atrial fibrillation, prior stroke, heart failure, and chronic renal insufficiency), Hunt-Hess grade,16 Fisher score,15 Glasgow Coma Scale score on admission,18 need for mechanical ventilation, cause of SAH, mode of treatment of aneurysm if applicable, secondary complications of SAH, length of stay, do-not-resuscitate orders, and disposition. Diagnoses of secondary complications of SAH were based on standardized definitions (Online Supplement; http://stroke.ahajournals.org).

Outcomes

The patients’ actual mRS scores at 6 months were determined by telephone interviews using a validated standardized script.19 The interviews were performed by a single neurovascular fellow blinded to previous prognostications. If severe disability precluded conversation with the patient, the patient’s closest surrogate was interviewed instead. For patients who we could not contact, we performed an online obituary search and contacted the patients’ primary care physicians to determine the patient’s vital status.

Statistical Analysis

Descriptive statistics were used to calculate the rates of good neurological outcomes, defined as an mRS score of ≤2 (ie, independent in activities of daily living), and poor outcomes, defined as an mRS score of 3 to 6 (ie, dependent in activities of daily living or dead). Our primary measure of interest was the accuracy of physicians in correctly predicting this dichotomized 6-month functional outcome. Accuracy was defined as the proportion of good and bad outcomes correctly predicted divided by the total number of predictions made and was calculated for the entire cohort. We also calculated the proportion of patients who had a good outcome when they were predicted to do well and the proportion of patients who had a poor outcome when predicted to do poorly. In addition, we evaluated the association between the specific predicted mRS and the actual 6-month mRS using a weighted χ² analysis, which accounted for the degree of discordance between predicted and actual mRS scores rather than penalizing all discordances equally. Each type of analysis was performed for each of the first 14 days of each patient’s hospitalization.

To assess for trends over the course of the hospitalization, we performed ordinal logistic regression on the physicians’ prognostications with the day of hospitalization as a predictor variable. To examine factors associated with the accuracy of physician prognostications, we used a logistic regression model accounting for clustering by patient with generalized estimating equations using an independence working correlation. To determine whether the timing of the physician prognostication affected its accuracy, we included in this model the day of prognostication as well as an interaction term between the prognostication and the day of prognostication. Also, to evaluate whether physician prognostication is useful independent of clinical factors, we included in our model the following clinical factors known from other studies to be associated with SAH outcome: age, Hunt-Hess grade, Fisher score, recurrent hemorrhage, infarction, and hydrocephalus.13,14,16,20

To account for the possibility of a self-fulfilling prophecy, in which the prediction of a poor outcome leads to withdrawal of care and thus an actual poor outcome, we performed a sensitivity analysis excluding patients for whom we recommended limitations in care and who died during their hospital stay. In addition, given the possibility of unequal prognostic accuracy among different fellows, we used Fisher exact test to compare the accuracy of individual fellows’ predictions on each day of hospitalization.

Statistical analyses were performed using Stata (Version 11; Stata Corporation, College Station, TX).

Results

Our cohort included 70 patients hospitalized with SAH; 66 (94%) had 6-month outcome data and were included in the final analysis. Patients’ demographic and clinical profiles were typical for patients with nontraumatic SAH (Table 1).13,14 The mean length of stay was 19 days (range, 2–101 days). Hospital disposition was most often to home (56%) followed by death (20%), acute rehabilitation (12%), subacute rehabilitation (6%), and other (6%); only 1 patient died outside of the hospital, and that patient had been discharged to a hospice facility in a vegetative state. The median mRS score at 6 months was 2 (interquartile range, 1–4) for the overall cohort and 2 (interquartile range, 1–2) for survivors. Forty patients (61%) had a good outcome and 26 (39%) had a poor outcome.

Study physicians made 648 total predictions during the first 14 days of patients’ hospitalization. More predictions were made early in the hospital course because some patients died or were discharged before Day 14 (eg, 92% of subjects remained alive and hospitalized on Day 3 compared with 39% on Day 14). Using a dichotomized categorization of
good (mRS 0–2) and bad (mRS 3–6) outcomes, the overall accuracy of prognostications ranged from 78% (95% CI, 65%–88%) on Day 2 to 88% (95% CI, 70%–98%) on Day 14 (Figure 1A). When physicians predicted that a patient would have a good outcome, the patient did well 81% (95% CI, 71% to 92%; Figure 1B) of the time; conversely, when physicians predicted that a patient would have a poor outcome, the patient did poorly 88% (95% CI, 77% to 99%; Figure 1C) of the time. There were 6 patients (9%) who physicians predicted would do poorly but who ultimately had a good outcome. In these patients, 20 of 77 predictions were incorrect (26%), although predictions were not far from actual outcomes at 6 months (ie, 1 patient was predicted to have an mRS of 4 but ultimately had an mRS of 2; 5 patients were predicted to have an mRS of 3 but ultimately had an mRS of 2 or 1).

A sensitivity analysis excluding 14 patients who died in the setting of care limitations did not demonstrate any clear worsening in overall accuracy as compared with the original cohort with overall accuracy ranging from 74% (95% CI, 61%–87%) to 87% (95% CI, 72%–100%; Figure 1D). However, the proportion of prognostications of a poor outcome that were borne out decreased to 74% (95% CI, 50%–98%). These incorrect prognostications of poor outcome involved 6 patients (12%).

Weighted analysis of the mRS score demonstrated moderate agreement between predicted and actual outcomes with values ranging from 0.43 (SE 0.10) on Day 12 to 0.56 (SE 0.09) on Day 1 with no clear trend toward improved diagnostic accuracy with time (Figure 2). Tabular representation of the physicians’ mRS score prognostications in relation to the actual mRS scores on the days with the highest and lowest weighted values further illustrates the degree of agreement between physician predictions and actual outcomes (Table 2).

There was no significant trend in the nature of physician predictions (P=0.82) or their accuracy (P=0.72) over the course of the hospitalization. The interaction term between physicians’ prognostication and the day of the prognostication was not significant, indicating that there was no effect of timing on physicians’ prognostic accuracy. In addition, physician prognostication remained a useful predictor of outcomes independent of clinical factors known to be associated with outcome in SAH. Before inclusion of such clinical factors in the model, each 1-point increase in the predicted mRS was associated with an OR of 0.22 for a good neurological outcome at 6 months. Inclusion of clinical factors did not significantly change this OR (0.16), suggesting that physician prediction is useful independent of known clinical predictors.

Age, infection, mechanical ventilation, seizures, and hydrocephalus were associated with prognostication accuracy (Table 3). Neither do-not-resuscitate status nor discordance in goals of care between the neurovascular team and the patient or surrogate appeared to affect physician accuracy. For most hospital days (12 of 14), there was no statistically significant difference in accuracy between fellows and therefore we did not cluster our primary analysis by individual fellows nor did we pursue further investigations of whether other characteristics of the prognosticating physician had an impact on prognostic accuracy.
Despite patients’ and families’ need for accurate prognosis,1–3 few data exist on the prognostic accuracy of physicians; this is especially true for neurologists, who are often asked for predictions of neurological outcomes.10,11 This is worrisome because the decision to withdraw life support may rest heavily on physicians’ prognostication.12,21 On the other hand, physicians are no longer viewed by the public as infallible; instead, one study found that 88% of surrogate decision-makers of critically ill patients doubt physicians’ abilities to prognosticate accurately.3

Two published reports have evaluated the prognostic abilities of neurologists and neurosurgeons. Despite the use of a simplified 3-tier functional outcome scale and a very experienced attending neurosurgeon, one found an unacceptable 44% rate of incorrect predictions, including overoptimistic predictions that could have led to the inappropriate withdrawal of care in patients with traumatic brain injury.11 The other demonstrated that neurointensivists correctly predict functional outcomes in neurologically critically ill patients in 80% of cases but also found that these physicians were generally overoptimistic with only 63% of patients predicted to do well actually having good outcomes.10 In addition, neither of these studies assessed the optimal timing of functional outcome predictions. Only 2 studies in adult medical–surgical intensive care unit patients have analyzed the accuracy of critical care physicians’ predictions of survival at different points in time.9,22 Poses et al22 found no difference in the reliability or discriminating power between prognostications made at admission and 48 hours later,
Table 2. Physicians’ Prognostications of Patients’ Modified Rankin Scale Scores at 6 Months Compared With Patients’ Actual Scores at 6 Months on the Days With the Highest (Day 1) and Lowest (Day 12) Weighted Kappa Values*  

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Predicted</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Day 12</th>
<th>Predicted</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Predictions within 1 of the actual value are in bold and italics. Weighted κ analysis is a statistical technique that measures interrater agreement by accounting for chance and the degree of discordance between raters (eg, predicted and actual modified Rankin Scale scores in our study).

whereas Meadow and colleagues9 demonstrated increased caretaker disagreement and no significant change in the predictive power of daily prognostications of death over time.

In our cohort, we found neurologists to be reasonably accurate at predicting dichotomous outcomes in patients with SAH with overall accuracy ranging from 78% to 88%. Physicians were slightly more accurate when predicting poor outcomes (88% actually did poorly) rather than good outcomes (81% actually did well). Overoptimistic errors by physicians are dangerous if they lead to aggressive care in situations when patients would have chosen comfort measures if given accurate prognostication, thus leading to patients living in severely disabled states that are incompatible with their previously known wishes. Conversely, an overpessimistic error could result in the inappropriate withdrawal of care in a patient. It is worrisome that our prognosticating physicians committed overpessimistic errors in 9% of our cohort. These errors highlight the fact that physician accuracy is imperfect and reinforce the message that physicians should explicitly mention uncertainty when prognosticating and should not make absolute predictions.1

Interestingly, the timing of our predictions had no impact on either the nature (optimistic versus pessimistic) or the accuracy of prognostications, supporting the work of prior authors.9,22 Therefore, the timing of discussions about goals of care may not be as important as transparency about the degree of physician uncertainty regarding outcomes. Given imperfect accuracy when predicting poor outcomes, physicians should be especially cautious about making poor outcome predictions early in patients’ hospital course.

The clinical variables associated with accuracy in our cohort are all factors that can cloud the neurological examination. Infections, seizures, and hydrocephalus all lead to impaired consciousness; mechanical ventilation requires sedation, which confounds the examination and can lead to a recrudescence of old deficits; and older patients often suffer from delirium in intensive care units.23,24 Physicians should be particularly cautious in making firm predictions in patients with these variables. Of note, neither do-not-resuscitate status nor discordance in goals of care between physicians and patients or surrogates affected physician accuracy.

Our results suggest that physician prognostication is a useful predictor of outcomes independent of objective clinical factors. Numerous studies have demonstrated that despite their shortcomings, physicians perform as well or better than objective prediction scores.5,7,8 Physicians can use relevant information not contained within a prediction score, may be more flexible in dealing with patient heterogeneity, and benefit from hard-to-define but real clinical intuition.2 However, no neurological clinical prediction score has been reported to combine physician intuition with objective clinical variables. Instead of only focusing on creating objective clinical prediction scores from large patient populations, it may be fruitful to also seek to combine such scores with physicians’ clinical intuition. This may be possible, for example, by identifying quantifiable factors that cloud physicians’ clinical intuition and then creating objective scores that can augment the physicians’ clinical judgment.

Our study has several important limitations. First, it was performed at a single, tertiary-care medical center with a high volume of SAH. Second, the prognosticating physicians were neurovascular fellows and not attendings; however, fellows at our institution are given significant autonomy and are responsible for reporting the teams’ outcome predictions to families. Nonetheless, neurologists with greater experience may have different prognostic accuracy than less seasoned physicians. Third, only the mRS score, which is heavily weighted by patients’ ability to ambulate, was used to assess prognostic accuracy. The use of more sophisticated scores incorporating neuropsychiatric and cognitive dysfunction, which are common long-term sequelae of SAH, could have altered our results. Fourth, we did not collect data on how or why prognosticating physicians made their determinations; such data could be instrumental in understanding the complex...
reasoning used by physicians to prognosticate. Fifth, without further validation, our results cannot necessarily be generalized to prognostications by vascular neurologists at community hospitals with low volumes of patients with SAH, nonvascular trained neurologists, or to patients with neurological diseases other than SAH. Sixth, because our study was initially designed as a quality assurance project, we did not perform an a priori power calculation and instead used a convenience sample; a larger sample size may have provided more precise estimates.

Neurologist predictions of outcomes after SAH are generally correct, but a concerning proportion (9%) of patients predicted to do poorly ultimately did well. Physicians should therefore explicitly discuss the uncertainty of their prognoses when communicating with families.

Disclosures
None.

References
Accuracy of Neurovascular Fellows' Prognostication of Outcome After Subarachnoid Hemorrhage
Babak B. Navi, Hooman Kamel, Charles E. McCulloch, Kazuma Nakagawa, Bharath Naravetla, Asma M. Moheet, Christine Wong, S. Claiborne Johnston, J. Claude Hemphill III and Wade S. Smith

*Stroke.* published online January 5, 2012;

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/early/2012/01/05/STROKEAHA.111.639161

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2012/01/05/STROKEAHA.111.639161.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to *Stroke* is online at:
http://stroke.ahajournals.org/subscriptions/
SUPPLEMENTAL MATERIAL

Supplemental Methods: Definition of secondary complications of subarachnoid hemorrhage.

Radiographic vasospasm was defined as a mean flow velocity > 120 cm/sec or Lindegaard ratio > 3 on transcranial Doppler ultrasound, or any documentation of vasospasm on the official report of computed tomogram angiography or catheter angiography. Clinically symptomatic vasospasm was defined as the development of new neurologic signs attributed by the treating physician to ischemia from vasospasm after other possible causes of neurologic worsening had been excluded. Infarction was defined as any new hypodensity on head computed tomogram or area of restricted diffusion on brain magnetic resonance imaging attributed by the interpreting neuroradiologist to ischemia (including infarcts related to procedures). Cerebral salt wasting was defined as a serum sodium level of <130 mEq/L in the setting of excessive natriuresis. Hydrocephalus was defined as ventriculomegaly out of proportion to age or the development of ventricular enlargement as compared to a previous image with an associated decline in neurologic function. Infection was defined as any potential bacterial or fungal infectious process leading to antibiotic therapy by the treating physicians; this broad clinical definition was chosen due to known complexities in identifying nosocomial infections in the neurocritical care population.

References: