Title

Evaluating Performance of the Spetzler-Martin Supplemented Model in Selecting Patients With Brain Arteriovenous Malformation for Surgery

Journal

Stroke, 43(9)

ISSN

0039-2499

Authors

Kim, Helen

Pourmohamad, Tony

Westbroek, Erick M

et al.

Publication Date

2012-09-01

DOI

10.1161/strokeaha.112.661942

Peer reviewed

Published in final edited form as:
Stroke. 2012 September ; 43(9): 2497–2499. doi:10.1161/STROKEAHA.112.661942.

Evaluating Performance of the Spetzler-Martin Supplemented Model in Selecting Brain Arteriovenous Malformation Patients for Surgery

Helen Kim, PhD[1,2], **Tony Pourmohamad, MA**[1], **Erick M. Westbroek, BS**[1], **Charles E. McCulloch, PhD**[1,2], **Michael T. Lawton, MD**[3], and **William L. Young, MD**[1,3,4]

[1]Center for Cerebrovascular Research, Department of Anesthesia and Perioperative Care, University of California, San Francisco

[2]Department of Epidemiology and Biostatistics, University of California, San Francisco

[3]Department of Neurological Surgery, University of California, San Francisco

[4]Department of Neurology, University of California, San Francisco

Abstract

Background—Our recently proposed point scoring model includes the widely-used Spetzler-Martin (SM)-5 variables, along with age, unruptured presentation, and diffuse border (SM-Supp). Here we evaluate the SM-Supp model performance compared to SM-5, SM-3, and Toronto prediction models using net reclassification index (NRI), which quantifies the correct movement in risk reclassification, and validate the model in an independent dataset.

Methods—Bad outcome was defined as worsening between preoperative and final postoperative modified Rankin Scale score. Point scores for each model were used as predictors in logistic regression, and predictions evaluated using NRI at varying thresholds (10–30%) and any threshold (continuous NRI>0). Performance was validated in an independent dataset (n=117).

Results—Net gain in risk reclassification was better using the SM-Supp model over a range of threshold values (NRI=9–25%) and significantly improved overall predictions for outcomes in the development dataset, yielding a continuous NRI of 64% versus SM-5, 67% versus SM-3, and 61% versus Toronto (all P<0.001). In the validation dataset, the SM-Supp model again correctly reclassified a greater proportion of patients versus SM-5 (82%), SM-3 (85%), and Toronto models (69%).

Conclusions—The SM-Supp model demonstrated better discrimination and risk reclassification than several existing models and should be considered for clinical practice to estimate surgical risk in BAVM patients.

Keywords

receiver operator curve; Modified Rankin Scale; net reclassification

Correspondence: Helen Kim, PhD, 1001 Potrero Avenue, Box 1363, San Francisco, CA 94110, Phone: 415-206-4789; Fax: 415-206-8907, kimhel@anesthesia.ucsf.edu.

Disclosures
None.

Introduction

The Spetzler-Martin (SM) 5-point grading scale is the most widely accepted surgical risk prediction tool for brain arteriovenous malformations (BAVM), although other models have been proposed.[1–6] We recently developed a simple point scoring model that incorporates SM angiographic variables but supplements with additional clinical factors (SM-Supp) to improve outcome prediction**,** and demonstrated improved discrimination over SM-5 using area under the receiver operating characteristic curve (AUROC).[7] Here we extend our previous work by comparing SM-Supp performance to other models using the net reclassification index (NRI), and validating the model in an independent dataset.

Methods

We included consecutive BAVM patients who underwent microsurgical resection between 2000–2010 with at least one post-operative visit and no missing outcome data. The *development dataset* consisted of 300 BAVM patients treated by a single neurosurgeon (MTL) between 2000 and 2007.[7] The primary *validation dataset* consisted of 117 patients (67 new MTL cases between 2007 and 2010; 50 cases from other neurosurgeons between 2000 and 2010) with no missing data. We also included data from a larger validation dataset (n=183) for which we multiply imputed missing angiographic data (provided in Online Supplement).

Outcome was change between pre-operative and last post-operative Modified Rankin Scale (MRS) score,[8] dichotomized into >0 (bad outcome) versus ≤0 (good outcome).[7] Predictors included age at surgery, sex, non-hemorrhagic presentation, AVM size, any deep venous drainage, eloquence, diffuse border, and time from surgery to last post-operative MRS assessment (days). SM-5,[1] SM-3,[6] Toronto[5] and SM-Supp[7] scores are defined in Supplemental Table S1.

NRI[9, 10] was used to evaluate model performance and quantifies the correct movement in risk reclassification when comparing predictions between two models at various risk thresholds (10–30%) or any threshold (continuous, cNRI>0).[10] NRI was compared by combining one-sided McNemar's tests across outcomes using Fisher's method.[11] We derived bootstrap 95% CI for cNRI using 1000 replications.

Results

Characteristics were similar between development and validation datasets (P>0.05, Table 1, Supplemental Table S2). Outcomes were bad for 73 (24%) and good for 227 (76%) patients in the development dataset. In the validation dataset, outcomes were bad for 39 (21%) and good for 144 (79%) patients.

In the development dataset, NRI showed improvement in reclassification of 9–25% with SM-Supp than SM-5 over all threshold values (Table 2). A greater net gain was observed at lower thresholds for good and at higher thresholds for bad outcomes. For example, at 15% risk threshold, 85 of 300 (28%) were reclassified into different risk categories. Net gain in reclassification was −6.8% for those with bad outcomes and 27% for those with good outcomes (NRI=0.205, P<0.001). Thus, patients with good outcomes were 21% more likely to move down a risk category than up, compared to patients with bad outcomes.

Because risk categories for BAVM surgical outcome are not well established, we also calculated the cNRI comparing SM-Supp to SM-5. The cNRI was 64% (95% CI=39–89%, P<0.001), with a net gain of 26% in those with good outcomes and 37% in those with bad outcomes (Table 2). Thus, 64% had predicted risks reclassified in the correct direction with

SM-Supp. Results were similar when comparing SM-Supp to SM-3 (cNRI=67%, 95% CI=41–93%) and to Toronto (cNRI=61%, 95% CI=37–85%). Scatterplots of predicted probabilities (Figure 1) by good and bad outcomes reflected a greater proportion of patients with correct assignments using the SM-Supp model compared to either SM-5 (Figure 1A), SM-3 (Figure 1B), or Toronto models (Figure 1C). In the validation dataset, the SM-Supp model again correctly re-classified a greater proportion of patients versus SM-5 (cNRI=82%, 95% CI=43.6–121%), SM-3 (cNRI=85%, 95% CI=44.7–126%), and Toronto models (cNRI=69%, 95% CI=26.4–121%).

Consistent with NRI results, the SM-Supp model yielded better discrimination and highest AUROC than all other models (Supplemental Figure S1) in development (AUROC=0.76, P<0.001) and validation (AUROC=0.77, P=0.402) datasets.

Discussion

The SM-Supp model performed equally well in predicting outcomes in an independent dataset, and consistently showed better risk reclassification and discrimination. For example, greater than 60% of patients were correctly reclassified as having higher risk for those with bad outcomes and lower risk for those with good outcomes compared to each of SM-5, SM-3, or Toronto models.

Direct comparisons with other models[2–5] are difficult because outcome measures and time points assessed differ among studies, e.g., we examined change in outcome, which takes into account pre-operative state. Only Spears et al[5] compared performance of their prediction model to SM-5 using mRS and AUROC, showing good discrimination and performance (AUROC=0.80).[5] Our model showed equally high discrimination in both development (AUROC=0.76) and validation datasets (AUROC=0.77).

Although the SM-Supp model derives from a single neurosurgeon and referral institution, we provide an independent validation using the NRI and include cases treated by other neurosurgeons in the largest series to date. However, further validation in external settings would be useful to assess generalizability and clinical utility. A limitation of all scoring systems is dealing with missing data. In our full validation dataset (n=183), 34% were missing angiographic data for SM-Supp, 36% for Toronto, and 13% for SM-5 and SM-3 scores. One way of accommodating missing data is through multiple imputation (see Online Supplement). Prospective studies planning to use SM-Supp should have minimal issues with missing data: all variables should be available from angiograms and MRI, which are standard for diagnostic evaluation and pre-treatment planning, or from records at clinic visits.

In conclusion, the SM-Supp model performs better than current prediction models, and should be considered for use in clinical practice. An online calculator is provided to assist clinicians (http://avm.ucsf.edu/healthcare_pro/).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Funding

Supported by K23NS058357 (HK), R01NS041877 (WLY), P01NS044155 (WLY), and Doris Duke Charitable Foundation (EMW).

References

1. Spetzler RF, Martin NA. A proposed grading system for arteriovenous malformations. J Neurosurg. 1986; 65:476–483. [PubMed: 3760956]
2. Tamaki N, Ehara K, Lin TK, Kuwamura K, Obora Y, Kanazawa Y, et al. Cerebral arteriovenous malformations: factors influencing the surgical difficulty and outcome. Neurosurgery. 1991; 29:856–861. [PubMed: 1758597]
3. Pertuiset B, Ancri D, Kinuta Y, Haisa T, Bordi L, Lin C, et al. Classification of supratentorial arteriovenous malformations. A score system for evaluation of operability and surgical strategy based on an analysis of 66 cases. Acta Neurochir (Wien). 1991; 110:6–16. [PubMed: 1882721]
4. Hollerhage HG, Dewenter KM, Dietz H. Grading of supratentorial arteriovenous malformations on the basis of multivariate analysis of prognostic factors. Acta Neurochir (Wien). 1992; 117:129–134. [PubMed: 1414513]
5. Spears J, Terbrugge KG, Moosavian M, Montanera W, Willinsky RA, Wallace MC, et al. A discriminative prediction model of neurological outcome for patients undergoing surgery of brain arteriovenous malformations. Stroke. 2006; 37:1457–1464. [PubMed: 16690897]
6. Spetzler RF, Ponce FA. A 3-tier classification of cerebral arteriovenous malformations. Clinical article. J Neurosurg. 2011; 114:842–849. [PubMed: 20932095]
7. Lawton MT, Kim H, McCulloch CE, Mikhak B, Young WL. A supplementary grading scale for selecting patients with brain arteriovenous malformations for surgery. Neurosurgery. 2010; 66:702–713. [PubMed: 20190666]
8. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. Stroke. 1988; 19:604–607. [PubMed: 3363593]
9. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. 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. [PubMed: 17569110]
10. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med. 2011; 30:11–21. [PubMed: 21204120]
11. Fisher, RA. Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd; 1932.

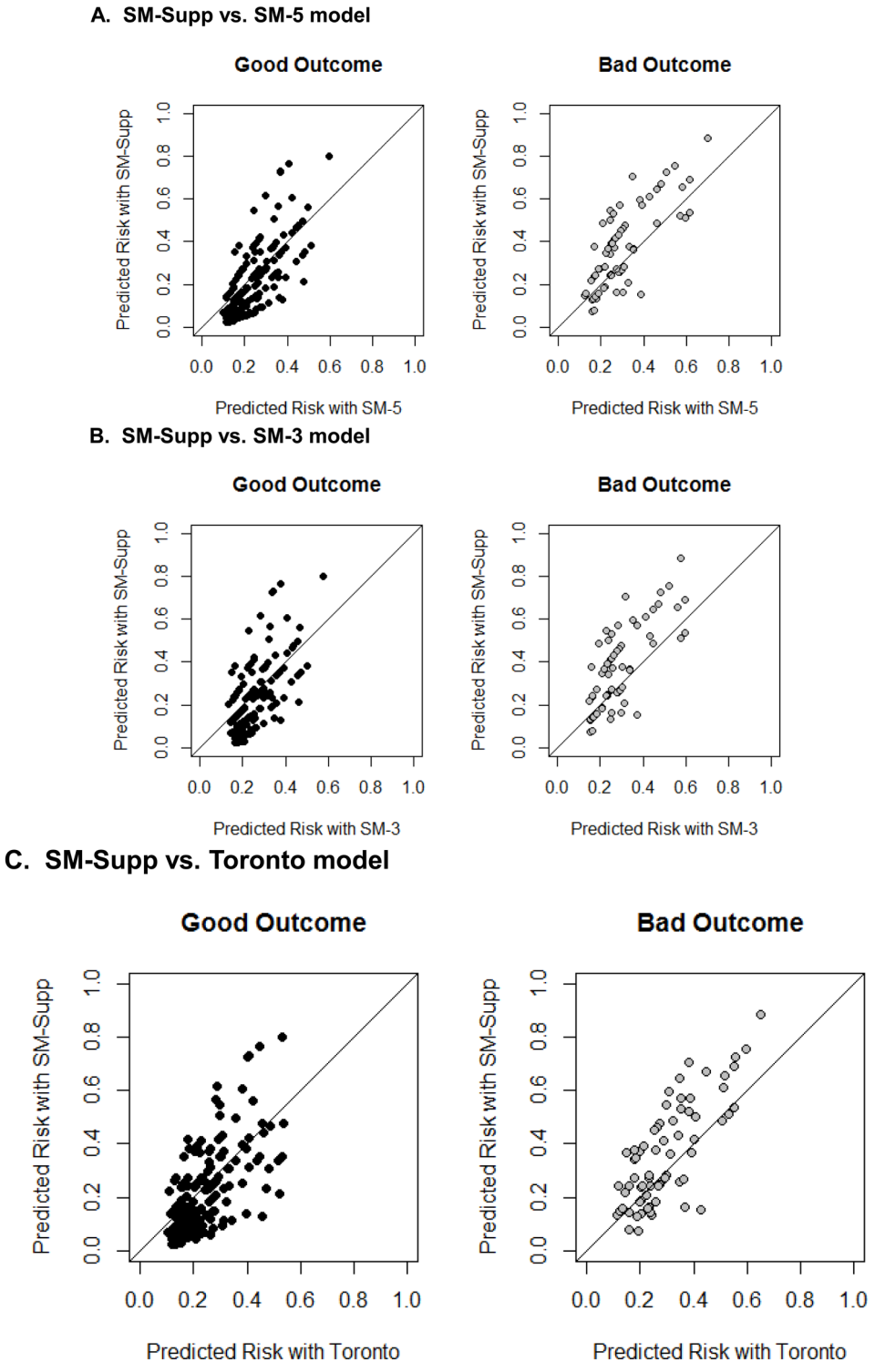

Figure 1. Scatterplot of predicted risk in patients with good (black dots) and bad (gray dots) post-surgical outcomes

The 45° line indicates concordance of predicted probabilities between models. For patients with good outcomes, a greater proportion of black dots were correctly assigned below the line indicating lower predicted risk using the SM-Supp model compared to either SM-5 (A),

SM-3 (B) or Toronto models (C). Conversely, in patients with bad outcomes, a greater proportion of gray dots were correctly classified above the line indicating higher predicted risk with SM-Supp.

Table 1

Pre-operative scores in the development and validation cohorts.

Scores	Development cohort n = 300	Validation cohort n= 117	P-value
Spetzler-Martin (SM-5)			
1	56 (19)	25 (21)	0.379
2	122 (40)	36 (31)	
3	91 (30)	43 (37)	
4	29 (10)	12 (10)	
5	2 (1)	1 (1)	
SM-Supplemented (SM-Supp)			
2	7 (2)	5 (4)	0.304
3	21 (7)	7 (6)	
4	55 (18)	28 (24)	
5	90 (30)	30 (26)	
6	70 (23)	32 (27)	
7	43 (15)	9 (8)	
8	9 (3)	3 (3)	
9	5 (2)	2 (2)	
10	0 (0)	1 (1)	
Modified Rankin Scale			
0	85 (28)	26 (22)	0.173
1	65 (22)	33 (28)	
2	33 (11)	21 (18)	
3	55 (18)	16 (14)	
4	33 (11)	9 (8)	
5	29 (10)	12 (10)	

Table 2

Net reclassification index (NRI) at varying risk thresholds and continuous NRI (>0) for improvement using Spetzler-Martin (SM) Supplemented versus SM-5 scale in development cohort.

Risk threshold	Bad Outcome Net gain	Good Outcome Net gain	NRI	P-value
10%	−0.027	0.278	0.250	<0.001
15%	−0.068	0.273	0.205	<0.001
20%	0.027	0.057	0.085	0.101
25%	0.096	0.044	0.140	0.031
30%	0.178	0	0.178	0.002
>0	0.260	0.374	0.635	<0.001