NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Validation of the Supplemented Spetzler-Martin Grading System for Brain Arteriovenous Malformations in a Multicenter Cohort of 1009 Surgical Patients

Helen Kim, PhD[2,3,4], **Adib A. Abla, MD**[1], **Jeffrey Nelson, MS**[2,4], **Charles E. McCulloch, PhD**[3,4], **David Bervini, MD**[7], **Michael K. Morgan, MD**[7], **Christopher Stapleton, MD**[6], **Brian P. Walcott, MD**[6], **Christopher S. Ogilvy, MD**[6], **Robert F. Spetzler, MD**[5], and **Michael T. Lawton, MD**[1,2,4]

[1]Departments of Neurological Surgery, University of California, San Francisco, CA

[2]Anesthesia and Perioperative Care, University of California, San Francisco, CA

[3]Epidemiology and Biostatistics, University of California, San Francisco, CA

[4]Center for Cerebrovascular Research, University of California, San Francisco, CA

[5]Division of Neurological Surgery, Barrow Neurological Institute, Phoenix, AZ

[6]Department of Neurological Surgery, Massachusetts General Hospital, Boston, MA

[7]Department of Neurological Surgery, Macquarie University, Sydney, Australia

Abstract

Background—The supplementary grading system for brain arteriovenous malformations (AVM) was introduced in 2010 as a tool for improving preoperative risk prediction and selecting surgical patients.

Objective—To demonstrate, in this multicenter validation study, that supplemented Spetzler-Martin grades have greater predictive accuracy than Spetzler-Martin grades alone.

Methods—Data collected from 1009 AVM patients who underwent AVM resection were used to compare predictive powers of Spetzler-Martin grades (SM) and supplemented Spetzler-Martin grades (SM-Supp). Patients included the original 300 UCSF patients plus those treated thereafter (N=117), and an additional 592 patients from three other centers.

Results—In the combined cohort, the SM-Supp system performed better than SM system alone: AUROC=0.75 (95% CI: 0.71 - 0.78) for SM-Supp and AUROC=0.69 (95% CI: 0.65 - 0.73) for SM (p < 0.001). Stratified analysis fitting models within three different follow-up groupings (< 6 months, 6 months – 2 years, and > 2 years) demonstrated that the SM-Supp system performed better than SM system for both medium (AUROC=0.71 vs. 0.62, p=0.003) and long follow-up

Correspondence: Michael T. Lawton, MD, Department of Neurological Surgery, University of California, San Francisco, 505 Parnassus Avenue, M780, Box 0112, San Francisco, CA 94143-0112, Phone: (415)353-3998; Fax: (415)353-3907, lawtonm@neurosurg.ucsf.edu.

(AUROC=0.69 vs. 0.58, p=0.001). Patients with SM-Supp grades   6 had acceptably low surgical risks (0 – 24%), with a significant increase in risk for grades above 6 (39% – 63%).

Conclusion—This study validates the predictive accuracy of the supplementary grading system in a multicenter cohort. SM-Supp grade of 6 is a cut-off or boundary for AVM operability. Supplemented grading is currently the best method of estimating neurological outcomes after AVM surgery, and we recommend it as a starting point in the evaluation of AVM operability.

Keywords

Arteriovenous malformation; microsurgery; patient selection; risk prediction; Spetzler-Martin grading system; supplementary grading system

INTRODUCTION

The supplementary grading system for brain arteriovenous malformations (AVM) was introduced in 2010 as a tool for improving preoperative risk prediction and selecting patients for surgery.[1] By analyzing additional factors outside of the Spetzler-Martin grading system that influence patient outcome after AVM resection, an analogous grading system was constructed that would supplement, rather than replace, the entrenched Spetzler-Martin system.[2] Points were assigned for the ABC's of AVMs: patient age, bleeding or hemorrhagic presentation, and AVM compactness, analogous to Spetzler-Martin scoring (Table 1). Pediatric patients (age < 20 years) were assigned 1 point; adults (age 20 – 40 years) were assigned 2 points; and older patients (> 40 years) were assigned 3 points. Patients presenting with unruptured AVMs were assigned 1 point and ruptured AVMs 0 points. Diffuse AVMs were assigned 1 point and compact AVMs 0 point. These points were added together for a supplementary AVM grade that ranged from 1 to 5. Simplicity is critical to the popularity of grading scales, and the supplementary grading scale was designed with this in mind. In addition, the two grading systems are analogous in their structure to make the supplementary grading scale memorable. In a consecutive surgical series of 300 patients, the supplementary grading system, sometimes referred to as the Lawton-Young grading system, had higher predictive accuracy than the Spetzler-Martin system and stratified surgical risk more evenly.[1] A sum of the two scores, or the supplemented Spetzler-Martin score, less than or equal to 6 identified patients with acceptably low surgical morbidity, providing a bedside numerical boundary for safe surgery.

This combination of two grading systems has proven to be a useful way of simplifying complex treatment decisions and has become an integralpart of our thinking about AVMs. However, the supplementary grading system was derived retrospectively from a single surgical series from one institution.[1] Validation of this grading system in a larger surgical series that includes other institutions and other neurosurgeons would encourage broader application. Therefore, we assembled such a cohort from 4 recognized centers with high-volume AVM practices and high-level neurosurgical expertise. In this report, we demonstrate that supplemented Spetzler-Martin scores have greater predictive accuracy than Spetzler-Martin scores alone, and that supplemented scoring is currently the best method of estimating neurological outcomes after AVM surgery.

METHODS

Study Population

The study was approved by the University of California, San Francisco Committee on Human Research and conducted in compliance with Health Insurance Portability and Accountability Act (HIPAA) regulations. Four institutions were invited to participate: UCSF, Barrow Neurological Institute (BNI; Phoenix, AZ), Massachusetts General Hospital (MGH; Boston, MA), and Macquarie University (MU; Sydney, Australia). The study sample consisted of patients with AVMs treated with microsurgical resection. UCSF patients included the original 300 patients plus those treated thereafter (N=117). Patients from outside institutions included those patients treated in the last 10 years with complete clinical, angiographic, and outcome data (BNI, N=266; MGH, N=153; and MU, N=173).

Study Variables

Data collected from 1,009 AVM patients who underwent microsurgical AVM resection was used to compare the predictive power of the Spetzler-Martin score (SM) and the supplemented Spetzler-Martin score (SM-Supp). Variables included the SM grade components (AVM size, venous drainage pattern, and eloquence),[2] the supplementary grade components (age at resection, hemorrhage prior to resection, and compactness of the AVM nidus),[1] pre-operative modified Rankin Scale (mRS) score, last available post-operativemRS score, and the time from resection to post-operativemRS. Data collection was harmonized using standard terminology.[3]

Statistical Analysis

Summary statistics of clinical characteristics were calculated by cohort. Differences amongst cohorts were assessed with chi-square tests for categorical characteristics and ANOVA for continuous characteristics. Outcomes were defined as the difference in mRS scores before and after surgery, and dichotomized into good (mRSunchanged or improved) or poor (mRSworse). The primary predictor variables were SM score (ranging from 1-5) and SM-Supp score (ranging from 2-10). Logistic regression models were adjusted further for log time from resection to last post-operativemRS assessment, because patients with longer follow-up have more time to recover and time may confound the relationship between mRSscore and outcome.

Performance of SM and SM-supp scores were evaluated by comparing the area under the receiver operating characteristic (AUROC) curves corresponding to SM and SM-Supp models, some times also referred to as c-statistics.[4] An AUROC of 0.5 indicates no discrimination whereas 1 indicates perfect discrimination. In general, an AUROC of 0.8 is considered clinically useful.[4] Data from all four institutions were combined in order to assemble one large cohort with a large sample size and tighter confidence intervals. To assess model fit, AUROC values were computed using 10-fold cross validation, which corrects for possible overfitting and inflated AUROC values.[5, 6] For this procedure, we partitioned the dataset into 10 mutually exclusive subsamples, each containing 10% of the data. Predicted values for subjects within a given subsample were calculated based on model fitting using the other 90% of the data.

Because timing of last mRS assessment is an important confounding variable that needed to be included in the logistic regression models, we performed a sensitivity analysis that removed the effect of time from the models. The data were subdivided into three groupings based on follow-up time (< 6 months, 6 months – 2 years, > 2 years) and SM and SM-Supp coefficients were calculated in each of these mutually exclusive groups. Given the lack of external validation data, 10-fold cross-validation was used for more honest assessments of the AUROC. p-values less than 0.05 were considered to be statistically significant. All analyses were performed using Stata/SE 13.1 software (StataCorp, College Station, TX).

RESULTS

Overall, 1009 patients were included in the combined cohort (Table 2). UCSF patients included 300 used for the original fitting of the SM-Supp model[1] and 117 used as an internal replication cohort.[7] Cohorts differed with respect to a number of AVM characteristics, including age at resection, distribution of AVM size, deep venous drainage, eloquence, unruptured presentation, and diffuse AVM nidus. The MGH cohort had a greater proportion of SM Grade I AVMs (41% vs. around 20% for other cohorts, p < 0.001). The UCSF cohort had a greater proportion of eloquent AVMs (p < 0.001). Compared to other cohorts, MU had a greater proportion of diffuse AVMs (40% vs. 18%, p < 0.001) and neurologically normal patients preoperatively (pre-operative mRS 0-1, 69% vs. 52% for other cohorts, p < 0.001), as well as the longest average follow-up time (2.5 ± 3.2 years vs. 1.7 ± 2.0 for other cohorts, p=0.004).

The predictive accuracy of the supplemented Spetzler-Martin scores was the same in the combined replication cohort of 709 patients from all 4 institutions as it was in the original cohort of 300 UCSF patients used to generate the SM-Supp model (AUROC 0.7482 vs. 0.7428, respectively, p=0.887; Figure 1). The UCSF internal replication cohort of 117 patients represents a prospective application of the supplemented SM system, and this supplemented SM system outperformed the SM system alone using a model fitting based on the original 300 UCSF patients (AUROC=0.77 vs. 0.71, respectively). Due to the small sample size, this difference was not statistically significant (p=0.102). However, when refitting the model in the combined cohort of 1009 patients, analysis of AUROC values for the SM and SM-Supp systems showed that the SM-Supp system still performed significantly better than the SM system alone: AUROC=0.69 (95% CI: 0.65 - 0.73) for SM and AUROC=0.75 (95% CI: 0.71 - 0.78) for SM-Supp (Figure 2). This difference was highly statistically significant (p < 0.001).

Stratified analysis fitting models within three different follow-up groupings (< 6 months, 6 months – 2 years, and > 2 years) demonstrated that the SM-Supp system performed better than SM system alone for both medium follow-up (AUROC=0.71 vs. 0.62, p=0.003) and long follow-up (AUROC=0.69 vs. 0.58, p=0.001). Differences were not significant for short follow-up (AUROC=0.70 vs. 0.67, p=0.21). Therefore, the SM-Supp system predicted poor outcomes more accurately than the SM system alone model beyond 6 months, independent of timing of last mRS assessment.

We explored adding other risk factors or changing the age cut-offs currently in the SM-Supp model. However, neither the addition of deep perforator supply (OR=0.88; 95% CI: 0.59-1.30; p=0.51) nor using alternative age cut-offs significantly increased the predictive power of the SM-Supp score or improved the grading system (data not shown).

Neurological outcomes were stratified by supplemented Spetzler-Martin grades, with surgical morbidity increasing with increasing grade (Table 3). SM-Supp Grade 6 was a boundary for AVM operability. Patients with supplemented Spetzler-Martin grades less than or equal to 6 had acceptably low surgical risks (0- 24%), with a significant increase in risk in patients with supplemented Spetzler-Martin grades greater than 6 (39% - 63%).

DISCUSSION

Validation of Supplemented Spetzler-Martin Grading System

Patient selection is the key to avoiding surgical complications and poor neurological outcomes with microsurgical resection of brain AVMs. The wide variety of AVM anatomy, size, location, and clinical presentation makes patient selection for surgery a difficult process. Neurosurgeons have analyzed their surgical experiences to identify factors that determine the risks of surgery in order to assist them in this selection process. Numerous classification schemes have been developed, each with its own emphasis, accuracy, advantages, and disadvantages.[8–13] These classification schemes have value because they transform complex decisions into algorithms. Some are complex and others simple, each striving to predict surgical risk and achieve bedside applicability. The Spetzler-Martin grading system continues to with stand the test of time and remains the predominant classification scheme.[2] However, it is crude and has deficiencies, like lumping the heterogeneous group of Grade III AVMs together without clarifying the management of subtypes.[14] The Spetzler-Martin system also has redundancies, with low-grade AVMs managed similarly with surgery and high-grade AVMs managed conservatively. Therefore, Spetzler and Ponce condensed the original 5-tier grading system into 3 tiers[15] and made broad treatment recommendations based on AVM class. Proponents of this simplification emphasize that the fewer classes correspond more directly with treatment recommendations. Opponents of this simplification argue that it does not simplify the analysis because the same scoring steps of the original Spetzler-Martin scale are required with an additional step to reclassify the AVM. Opponents also emphasize that the class-specific recommendations are vague and encumbered with exceptions. For example, the class system still does not shed light on the heterogeneous grade III lesions. Patient selection is a sophisticated process that requires more complexity, not less, which is why the supplementary grading system was proposed.

This study validates the supplemented Spetzler-Martin grading system in a large, multi center cohort of over one thousand surgical patients. The AUROC value for the SM-Supp system was better than for the SM system alone (AUROC=0.75 vs 0.69, p < 0.001), (Figure 2). Prospective application of the supplemented Spetzler-Martin system in a small subgroup of patients resulted in an even higher AUROC value (AUROC = 0.77),[7] indicating improved predictive accuracy when the system becomes part of the decision-making process. This study also confirmed that a supplemented Spetzler-Martin grade of 6 was a cut-off or

boundary for AVM operability. Supplemented Spetzler-Martin grading embodies 6 of the most important surgical variables, stratifies AVMs across 9 possible grades, and has one clear threshold value, yet still remains simple, memorable, and applicable at the bedside in our clinical experience. Statistical validation showing improved predictive accuracy in a large cohort should encourage others to adopt this augmented grading system.

Elements of Supplementary Grading

The supplementary grading system works because patient age, hemorrhagic presentation, and compactness have important surgical consequences. Youth is associated with tolerance to surgery, recoverability, and neural plasticity, and is therefore weighted heavily in the supplementary grading system. AVM hemorrhage separates portions of the nidus from adjacent brain, thereby performing some parenchymal dissection. Hematoma evacuation immediately accesses this plane as if it were a free cortical surface and opens working room. Hematomas often extend to the brain surface and localize AVMs that are not cortically based. Hematomas lead to deep AVMs without a superficial draining vein, and clot evacuation opens non-anatomical corridors of exposure that do not exist in patients with unruptured AVMs. Hematoma above nidus opens a transcortical corridor to its top side, while hematoma adjacent to nidus dissects its parenchymal sides and hematoma beneath nidus facilitates deep dissection. Accessing an AVM through the hematoma cavity requires no corticectomy, produces no additional morbidity, and reaches AVMs in deep lobar white matter, basal ganglia, and thalamus. Absorption of hematoma after rupture reduces, but does not eliminate, these surgical advantages. Blood is replaced with chronic inflammation, scar tissue, gliosis, and encephalomalacia. Old hemorrhage cavities can be exploited as routes of AVM access, and thickened gliotic planes handle better and separate more cleanly than soft, friable, unbled planes.

A compact AVM is wound tightly with distinct margins, no intermixed brain, and separability from adjacent brain, all of which clarify parenchymal dissection. In contrast, a diffuse AVM is wound loosely, as if pulled apart or unraveled, with indistinct margins, intermixed brain, and poor separability, all of which complicate parenchymal dissection. Diffuseness comes in many variations: a lacey web of small arterioles; tentacles of large arteries; a ragged lattice of arteries and arterioles spread across a pial surface; a network of leptomeningeal collateral arteries in a watershed zone between vascular territories; or wisps of perforating arteries penetrating the nidus from below. Compact AVMs are easily deciphered and followed, whereas diffuse AVMs force the neurosurgeon to establish the plane of separation between AVM and brain. The margin of a diffuse AVM can be an unruly fringe that must be encircled either widely at the expense of intervening brain tissue, or closely in a way that crosses more of this arterial fringe. The interplay of eloquence and hemostasis pull in and push back the circum dissection, challenging the neurosurgeon to find the right dissection distance. In some cases, this judgment can bring the dissection too close to the nidus and cause bleeding or leave behind abnormal portion of nidus. In other cases, this judgment can bring the dissection too far from the nidus, resulting in complete and easier AVM resection but removing more brain tissue than is necessary. The difficulties in circum dissecting diffuse AVMs can lead to tissue injury, parenchymal bleeding, contusions,

postoperative edema, seizures, and delayed hemorrhagic complications, all of which impact patient outcomes and account for its place in the supplementary grading system.

Limitations

This study is limited by selection and inclusion biases. Operated patients were highly selected by the participating neurosurgeons and selection criteria were not uniform across centers. Only those patients with complete clinical and radiographic data necessary for grading and outcome assessment were included, resulting in an unknown number of excluded patients. Participating neurosurgeons have different surgical techniques, technical skills, and experience, but each neurosurgeon directs a busy AVM service, has experience with over 500 AVM operations, and has a high level of surgical expertise. This study was a retrospective review of clinical and radiographic data, with the exception of the 117 patients from UCSF analyzed prospectively after creating the supplementary grading system.[7] However, each of these centers maintains an active vascular registry and much of the data in these other 892 patients were collected prospectively.

Scoring diffuseness is a limitation of the supplementary grading system because it is difficult to quantify and may vary between observers. In fact, diffuseness scores varied between 12% and 40% at different centers. We developed methods using digital angiography, contour plots, and area-intensity profiles to quantitate diffuseness.[16] Computer algorithms generate outlines around AVMs at set threshold image intensities, which are converted to AVM areas.[16] As the threshold intensity is dialed down, the area of the contour plot increases linearly with compact AVMs, but exponentially with diffuse AVMs. In other words, as the AVM is outlined more generously, compact AVMs grow a little and diffuse AVMs grow a lot. Area-intensity profiling is a cumbersome research tool for defining diffuseness and is not applicable at the bedside. Instead, as a general definition, a circle can be drawn easily around a compact AVM on an angiogram, but not so easily around a diffuse one. The subjectivity of diffuseness should not taint the supplementary grading system. Eloquence is also somewhat subjective and imprecise, given the known translocations of neurological function associated with AVMs, and this has not tainted the Spetzler- Martin grading system. In fact, eloquence scores also varied in this study between 28% and 60% at different centers – similar to the range in diffuseness scores.

Although critics might suggest that an AUROC value of 0.75 is low and question the predictive power of the supplemented Spetzler-Martin grading system, this value exceeds that of the Spetzler-Martin grading system (0.69), which remains the gold standard to which all other grading systems are compared. Establishing a cutoff for AVM operability requires an evaluation of the AVM's natural history, the morbidity of therapeutic alternatives, the patient's age and co-morbidities, and his or her unique emotional state, including treatment preferences, expectations, risk tolerance, and anxieties. These factors influence the cutoff for AVM operability in each individual patient and make it difficult to establish a universal cutoff. Although it may seem arbitrary in this context, our cutoff score of 6 reflects a significant increase in morbidity observed in patients with supplemented scores greater than 6, as well as a risk (0% – 24%) that many AVM patients might find acceptable in pursuit of

AVM cure. Our stated cutoff should be viewed as a guideline and we encourage individualizing AVM treatment recommendations.

CONCLUSION

This study demonstrates the predictive accuracy of the supplementary grading system in a large, multicenter cohort and validates the use of supplemented Spetzler-Martin scores. Were commend it as a starting point in the evaluation of AVM operability. No grading system, combination of grading systems, or simple algorithm can replace the discriminating process of patient selection. We offer this supplementary grading system as just another tool to guide the process of analyzing some of the critical factors that influence patient outcome, in order to make more rational choices when weighing known risk of spontaneous AVM rupture against risk of intervention. The supplementary grading system is intended to improve preoperative risk prediction, and we expect that it will assist in patient selection for surgery. The current crop of grading systems is generally useful, but imperfect and evolving. As the pathophysiology, hemodynamics, and genetics of AVMs are elucidated through research, grading systems will incorporate these advances. Therefore, the work of developing AVM grading systems should be viewed as an ongoing process, and clinicians should be open to revising established, proven grading systems.

Acknowledgment

REFERENCES

1. 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 Apr; 66(4):702–713. discussion 713. [PubMed: 20190666]

2. Spetzler RF, Martin NA. A proposed grading system for arteriovenous malformations. Journal of neurosurgery. 1986 Oct; 65(4):476–483. [PubMed: 3760956]

3. Atkinson RP, Awad IA, Batjer HH, et al. Reporting terminology for brain arteriovenous malformation clinical and radiographic features for use in clinical trials. Stroke; a journal of cerebral circulation. 2001 Jun; 32(6):1430–1442.

4. Harrell, FE. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Springer; 2001.

5. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Paper presented at: IJCAI1995;

6. Westerhuis JA, Hoefsloot HC, Smit S, et al. Assessment of PLSDA cross validation. Metabolomics. 2008; 4(1):81–89.

7. Kim H, Pourmohamad T, Westbroek EM, McCulloch CE, Lawton MT, Young WL. Evaluating performance of the spetzler-martin supplemented model in selecting patients with brain arteriovenous malformation for surgery. Stroke; a journal of cerebral circulation. 2012 Sep; 43(9): 2497–2499.

8. Davies JM, Kim H, Young WL, Lawton MT. Classification schemes for arteriovenous malformations. Neurosurgery clinics of North America. 2012 Jan; 23(1):43–53. [PubMed: 22107857]

9. Griessenauer CJ, Miller JH, Agee BS, et al. Observer reliability of arteriovenous malformations grading scales using current imaging modalities. Journal of neurosurgery. 2014 May; 120(5):1179–1187. [PubMed: 24628617]

10. Hollerhage HG, Dewenter KM, Dietz H. Grading of supratentorial arteriovenous malformations on the basis of multivariate analysis of prognostic factors. Acta neurochirurgica. 1992; 117(3–4):129–134. [PubMed: 1414513]

11. Pertuiset B, Ancri D, Kinuta Y, 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 neurochirurgica. 1991; 110(1–2):6–16. [PubMed: 1882721]

12. Spears J, Terbrugge KG, Moosavian M, et al. A discriminative prediction model of neurological outcome for patients undergoing surgery of brain arteriovenous malformations. Strokel a journal of cerebral circulation. 2006 Jun; 37(6):1457–1464.

13. Tamaki N, Ehara K, Lin TK, et al. Cerebral arteriovenous malformations: factors influencing the surgical difficulty and outcome. Neurosurgery. 1991 Dec; 29(6):856–861. discussion 861–853. [PubMed: 1758597]

14. Lawton MT. Spetzler-Martin Grade III arteriovenous malformations: surgical results and a modification of the grading scale. Neurosurgery. 2003 Apr; 52(4):740–748. discussion 748–749. [PubMed: 12657169]

15. Spetzler RF, Ponce FA. A 3-tier classification of cerebral arteriovenous malformations. Clinical article. Journal of neurosurgery. 2011 Mar; 114(3):842–849. [PubMed: 20932095]

16. Du R, Keyoung HM, Dowd CF, Young WL, Lawton MT. The effects of diffuseness and deep perforating artery supply on outcomes after microsurgical resection of brain arteriovenous malformations. Neurosurgery. 2007 Apr; 60(4):638–646. discussion 646–638. [PubMed: 17415200]
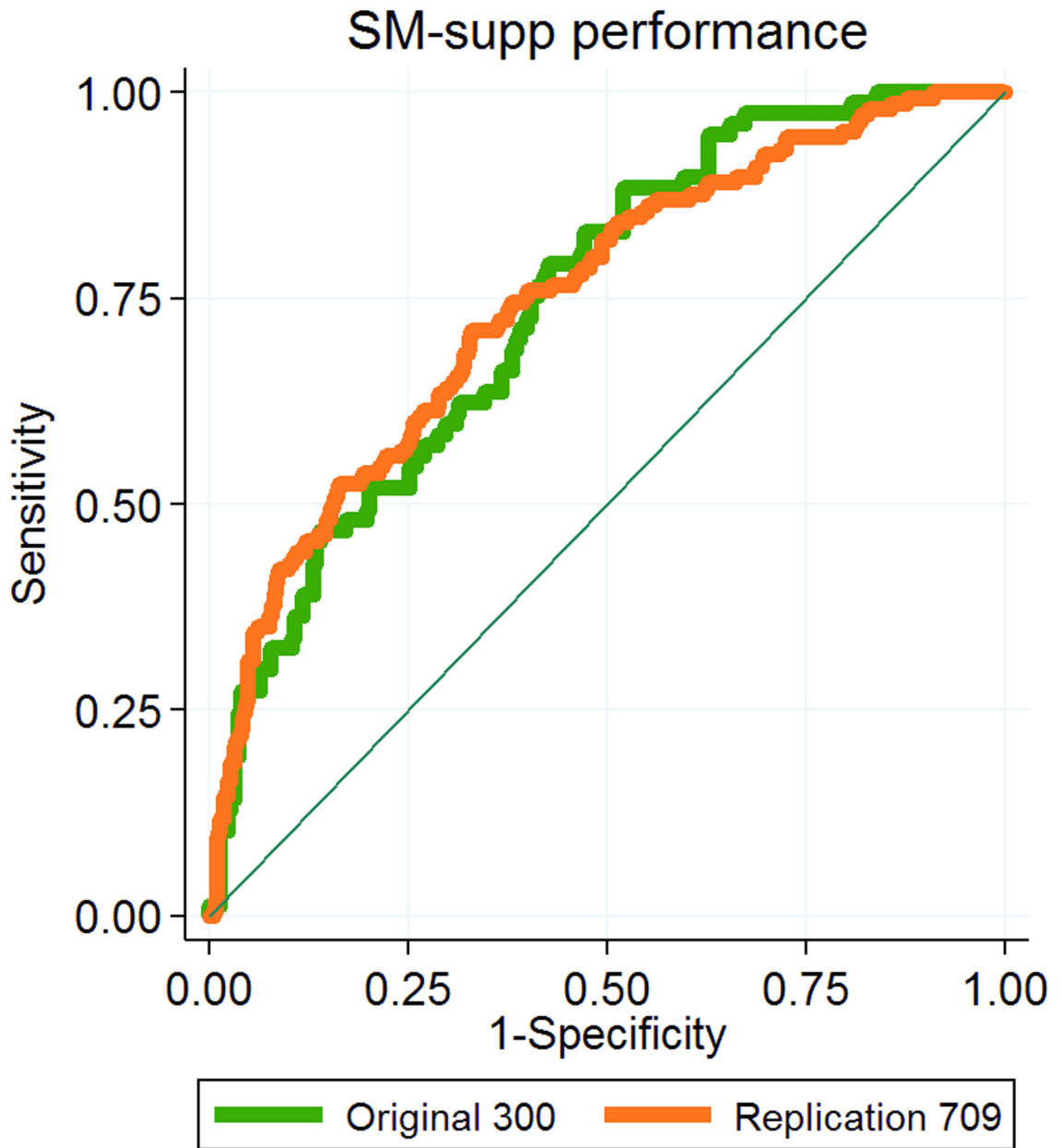
Figure 1.
Graph showing receiver operating characteristic (ROC) analyses for the Supplemented Spetzler-Martin grading system in the combined replication cohort of 709 patients from all 4 institutions (orange curve) and the original cohort of 300 UCSF patients used to generate the SM-Supp model (green curve). There were no significant differences in predictive accuracy in these two groups (AUROC 0.7482 vs. 0.7428, respectively, p=0.887).
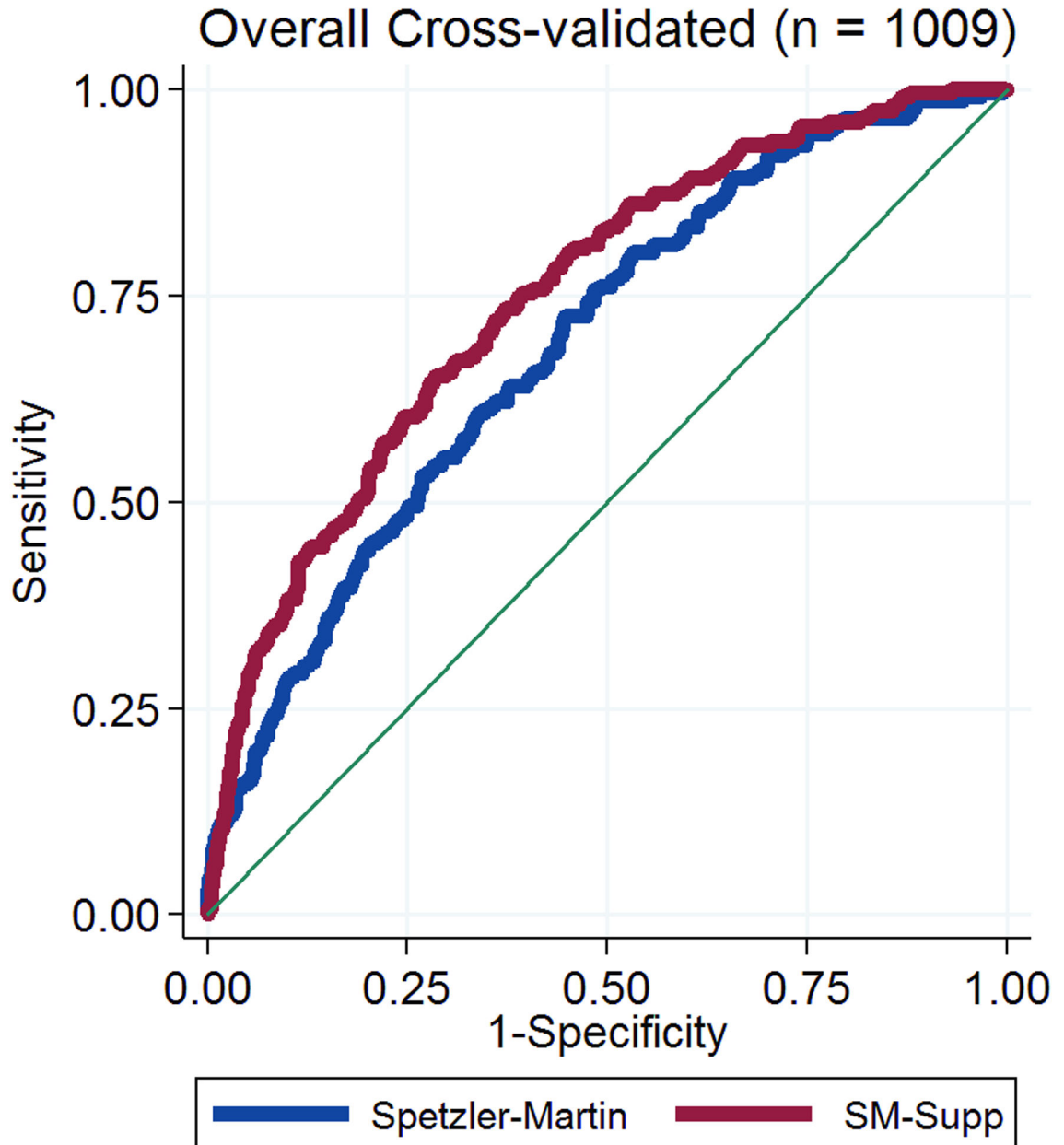
Figure 2.
Graph showing receiver operating characteristic (ROC) analyses for the Spetzler-Martin grading system (blue curve) and the Supplemented Spetzler-Martin grading system (red curve) in the combined cohort of 1009 patients. The predictive accuracy of the Supplemented Spetzler-Martin grading system was greater than that of the Spetzler-Martin grading system: AUROC = 0.75 vs 0.69, respectively; p < 0.001.

Table 1

Comparison of Spetzler-Martin and Supplementary grading systems.

Spetzler-Martin Grading		Points	Supplementary Grading	
Size	< 3 cm	1	Age	< 20 years
	3 – 6 cm	2		20 – 40 years
	> 6 cm	3		> 40 years
Venous Drainage	Superficial	0	Bleeding	Yes
	Deep	1		No
Eloquence	No	0	Compactness	Yes
	Yes	1		No
Total		5		

Table 2

Comparison of patients and AVM characteristics in cohorts at 4 institutions.

	UCSF Retrospective n=300	UCSF Prospective n=117	BNI n=266	MGH n=153	MU n=173	p-value
Age at resection (y)						<0.001
<20	47 (16%)	33 (28%)	61 (23%)	10 (7%)	29 (17%)	
20–40	113 (38%)	29 (25%)	91 (34%)	44 (29%)	74 (43%)	
>40	140 (47%)	55 (47%)	114 (43%)	99 (65%)	70 (40%)	
AVM Size (cm)						<0.001
<3	194 (65%)	75 (64%)	154 (58%)	110 (72%)	63 (36%)	
3–6	101 (34%)	40 (34%)	105 (39%)	42 (27%)	102 (59%)	
>6	5 (2%)	2 (2%)	7 (3%)	1 (1%)	8 (4%)	
Deep venous drainage	130 (43%)	48 (41%)	107 (40%)	40 (26%)	62 (36%)	0.007
Eloquence	160 (53%)	70 (60%)	118 (44%)	43 (28%)	73 (42%)	<0.001
Non-hemorrhagic presentation	143 (48%)	48 (41%)	138 (52%)	81 (53%)	112 (65%)	0.001
Diffuse nidus	37 (12%)	15 (13%)	59 (22%)	42 (27%)	69 (40%)	<0.001
Spetzler-Martin Grade						
1	55 (18%)	25 (21%)	64 (24%)	63 (41%)	28 (16%)	
2	122 (41%)	36 (31%)	92 (35%)	56 (37%)	69 (40%)	
3	92 (31%)	43 (37%)	79 (30%)	31 (20%)	48 (28%)	
4	29 (10%)	12 (10%)	30 (11%)	3 (2%)	24 (14%)	
5	2 (1%)	1 (1%)	1 (<1%)	0 (0%)	6 (2%)	
Supplemented Spetzler-Martin Grade						0.001
2	8 (3%)	5 (4%)	8 (3%)	2 (1%)	1 (1%)	
3	20 (7%)	8 (7%)	23 (9%)	5 (3%)	4 (2%)	
4	55 (18%)	27 (23%)	54 (20%)	35 (23%)	32 (19%)	
5	92 (31%)	31 (27%)	66 (25%)	63 (41%)	39 (23%)	
6	72 (24%)	31 (27%)	65 (24%)	19 (12%)	45 (26%)	
7	41 (14%)	9 (8%)	32 (12%)	23 (15%)	33 (19%)	
8	9 (3%)	3 (3%)	15 (6%)	5 (3%)	16 (9%)	
9	3 (1%)	2 (2%)	3 (1%)	1 (1%)	2 (1%)	

	UCSF Retrospective n=300	UCSF Prospective n=117	BNI n=266	MGH n=153	MU n=173	p-value
10	0 (0%)	1 (1%)	0 (0%)	0 (0%)	1 (1%)	
Pre-op mRS						<0.001
0	85 (28%)	26 (22%)	41 (15%)	22 (14%)	94 (54%)	
1	65 (22%)	33 (28%)	109 (41%)	53 (35%)	26 (15%)	
2	33 (11%)	21 (18%)	60 (23%)	39 (25%)	25 (14%)	
3	55 (18%)	16 (14%)	26 (10%)	17 (11%)	13 (8%)	
4	33 (11%)	9 (8%)	12 (5%)	13 (9%)	3 (2%)	
5	29 (10%)	12 (10%)	18 (7%)	9 (6%)	12 (7%)	
Post-op mRS						<0.001
0	96 (32%)	43 (37%)	91 (34%)	28 (18%)	99 (57%)	
1	88 (29%)	34 (29%)	74 (28%)	54 (35%)	27 (16%)	
2	52 (17%)	25 (21%)	54 (20%)	45 (29%)	33 (19%)	
3	21 (7%)	9 (8%)	22 (8%)	16 (10%)	6 (3%)	
4	15 (5%)	5 (4%)	13 (5%)	3 (2%)	0 (0%)	
5	1 (<1%)	0 (0%)	6 (2%)	2 (1%)	1 (1%)	
6	27 (9%)	1 (1%)	6 (2%)	5 (3%)	7 (4%)	
Worsening of mRS	77 (26%)	25 (21%)	51 (19%)	35 (23%)	34 (20%)	
Time from surgery to last mRS (y)	1.69 ± 1.69	0.99 ± 1.05	2.12 ± 2.35	1.80 ± 2.42	2.51 ± 3.17	0.008*

Values are the number observed with the specified value (and percentage) or the mean ± standard deviation.

P-values are derived from a chi-square test or ANOVA.

This p-value is based on comparison of log-transformed values.

Table 3

Neurological outcomes associated with Spetzler-Martin grades, Supplementary grades, and Supplemented Spetzler-Martin grades.

	Neurological Outcome			
	Worse		**Improved, Unchanged**	
	N	**%**	**N**	**%**
Spetzler-Martin Grade				
I	17	7%	218	93%
II	80	21%	295	79%
III	84	29%	209	71%
IV	37	38%	61	62%
V	4	50%	4	50%
Supplementary Grade				
I	5	6%	84	94%
II	30	16%	161	84%
III	70	18%	314	82%
IV	92	32%	196	68%
V	25	44%	32	56%
Supplemented Spetzler-Martin Grade				
2	0	0%	24	100%
3	1	2%	59	98%
4	21	10%	182	90%
5	54	19%	237	81%
6	56	24%	176	76%
7	54	39%	84	61%
8	30	63%	18	38%
9	6	55%	5	45%
10	0	0%	2	100%