Intrarater and interrater reliability of the pediatric arteriovenous malformation compactness score in children

Clinical article

FABIO A. FRISOLI, B.S.,1,2 SHIH-SHAN LANG, M.D.,1,2 ARASTOO VOSSOUGH, M.D., PH.D.,3 ANNE MARIE CAHILL, M.B.B.CH.,1 GREGORY G. HEUER, M.D., PH.D.,1,2 HISHAM M. DAHMOSH, M.B.B.CH.,3 PHILLIP B. STORM, M.D.,1,2 AND LAUREN A. BESLOW, M.D., M.S.C.E.4

1Department of Neurosurgery, University of Pennsylvania Medical Center, Philadelphia; 2Division of Neurosurgery and 3Department of Radiology, Children’s Hospital of Philadelphia, Pennsylvania; and 4Departments of Pediatrics and Neurology, Yale University School of Medicine, New Haven, Connecticut

Object. Cerebral arteriovenous malformations (AVMs) have a higher postresection recurrence rate in children than in adults. The authors’ previous study demonstrated that a diffuse AVM (low compactness score) predicts postresection recurrence. The aims of this study were to evaluate the intra- and interrater reliability of the AVM compactness score.

Methods. Angiograms of 24 patients assigned a preoperative compactness score (scale of 1–3; 1 = most diffuse, 3 = most compact) in the authors’ previous study were rerated by the same pediatric neuroradiologist 9 months later. A pediatric neurosurgeon, pediatric neuroradiology fellow, and interventional radiologist blinded to each other’s ratings, the original ratings, and AVM recurrence also rated each AVM’s compactness. Intrarater and interrater reliability were calculated using the \( \kappa \) statistic.

Results. Of the 24 AVMs, scores by the original neuroradiologist were 1 in 6 patients, 2 in 16 patients, and 3 in 2 patients. Intrarater reliability was 1.0. The \( \kappa \) statistic among the 4 raters was 0.69 (95% CI 0.44–0.89), which indicates substantial reliability. The interrater reliability between the neuroradiologist and neuroradiology fellow was moderate (\( \kappa = 0.59 \) [95%CI 0.20–0.89]) and was substantial between the neuroradiologist and neurosurgeon (\( \kappa = 0.74 \) [95%CI 0.41–1.0]). The neuroradiologist and interventional radiologist had perfect agreement (\( \kappa = 1.0 \)).

Conclusions. Intrarater and interrater reliability of the AVM compactness score were excellent and substantial, respectively. These results demonstrate that the AVM compactness score is reproducible. However, the neuroradiologist and interventional radiologist had perfect agreement, which indicates that the compactness score is applied most accurately by those with extensive angiography experience.

(https://thejns.org/doi/abs/10.3171/2013.2.PEDS12465)

Key Words • arteriovenous malformation • surgery • angiogram • recurrence • pediatric neurosurgery • compactness • reliability • vascular disorders

Cerebral AVMs in the pediatric population follow a distinct natural history that is different from that in adults. Arteriovenous malformations in children carry an increased yearly risk of hemorrhage with resultant significant morbidity and mortality.1,2 Consequently, AVMs are often treated aggressively with microsurgical resection, endovascular embolization, radiosurgery, or a combination of these treatment modalities.5,12 While the true postoperative recurrence rate of cerebral AVMs in the pediatric population has not been studied extensively, the literature suggests that it is higher than that
in adults.\(^1\)\(^,\)\(^7\) Many centers do not follow in adult patients
with serial vascular imaging after a postoperative image
has demonstrated an absence of residual AVM. It is pos-
tulated that higher postoperative recurrence in children is
due in large part to active angiogenesis of the immature
cerebral vasculature.\(^6\)\(^,\)\(^8\)\(^,\)\(^9\)

Certain characteristics of the AVM also likely con-
tribute to a higher recurrence risk. The Spetzler-Martin
grating scale, which considers AVM nidus size, location
in eloquent cortex, and venous drainage, has been widely
used to predict the incidence of postoperative neurologi-
cal complications.\(^1\)\(^,\)\(^3\) A large cohort study of mostly adult
patients with AVMs previously demonstrated that dif-
fuseness of the nidus and deep perforating arterial supply
are 2 additional characteristics that predict worse surgi-
cal outcome. Patients with noncompact nodules were 38%\ less likely to have overall improvement in preoperative
symptoms and were 5 times more likely to have postop-
erative clinical deterioration.\(^3\) Our recent data on pediat-
ric patients who underwent resection of cerebral AVMs
demonstrated that AVM compactness predicts AVM re-
currence after resection.\(^1\)\(^,\)\(^1\)\(^1\) Of 24 pediatric patients who
were assigned a pretreatment angiographic compactness
score by a neuroradiologist (scale of 1–3, 1 designating
the most diffuse), all patients with recurrence had a score
of 1 (\(p = 0.0003\)). This represents the first documenta-
tion in the literature of a compactness score for cerebral
AVMs in a pediatric population. The compactness score
was graded by a single neuroradiologist at one time point;
thus, the reproducibility of the score among raters was not
determined.

The aims of the current study were to determine the
intra- and interrater reliability of angiographic grading
of the compactness score for pediatric AVMs. The score
may be a helpful tool for radiologists and neurosurgeons
in predicting recurrence of cerebral AVMs in children
after resection. Such a score could eventually lead to a re-
duction in the use of serial angiograms in some children
at low risk for a recurrent AVM, a reduction that would
limit exposure to radiation and sedation.

**Methods**

**Study Protocol**

In our previous study, 24 pediatric patients with
angiographic evidence of an AVM were assigned a pre-
operative AVM compactness score visually by a board-
certified pediatric neuroradiologist.\(^1\)\(^,\)\(^1\)\(^1\) This noncom-
puterized compactness score was graded on a scale of 1 to
3 and was different from the computer-aided method of
determining nidus compactness described by Du et al.\(^3\)
A score of 1 designates a diffuse AVM, and a score of
3 designates a compact AVM. For the current study, all
clinically obtained preoperative vascular studies includ-
ing conventional angiography, CT angiography, and
MRI/MR angiography were used. For assessment of in-
trarater reliability, the same neuroradiologist rescored the
24 AVMs 9 months after the original scoring had been
performed.

Three new raters, a pediatric neurosurgeon, a pedi-
atriec intervention, and a pediatric neurora-
dology fellow, were then trained by the principal neu-
radiologist rater to score AVM compactness. The rat-
ers attended a training session with the principal rater to
review the definitions of a compactness score of 1 (most
diffuse), 2, and 3 (most compact) (Table 1; Figs. 1–3). Of
note, a diffuse AVM (compactness score of 1) may be
large or small, and the rating system is based on the over-
all compactness of the nidus and perinidal anomalous
vessels. Additionally, it is very important to assess the
images closely for very small and narrow-caliber abnormal
vessels that may not be evident immediately on a gross
overview of the images. The presence of these very small,
narrow-caliber abnormal vessels with intervening brain
tissue indicates a diffuse AVM that should be scored as
1. The raters scored 8 practice images from patients who
were not included in this study and reviewed the scoring
as a group to calibrate the ratings. The raters scored all
24 patients’ AVMs in the same order to limit variability
for each patient due to rater experience. The raters were
blinded to the other raters’ scores, to the original scores,
and to postoperative recurrence. The institutional review
board of the Children’s Hospital of Philadelphia approved
the study.

**Statistical Analysis**

Analyses were conducted using STATA software (ver-
sion 11.1, STATA Corp.). For the purposes of this study,
the original neuroradiologist’s scoring was used as the gold
standard for reliability calculations. Intrarater and inter-
rater reliability of the AVM compactness scores were as-
essed using the parameter kappa (\(k\)). Intrarater reliability
was assessed across all 4 raters and also between each new
rater and the principal rater. The \(k\) statistic was consid-
ered moderate agreement if it was between 0.41 and 0.60,
substantial agreement if between 0.61 and 0.80, and almost
perfect (excellent) if between 0.81 and 1.00.\(^10\) A 2-sided
probability value of \(\leq 0.05\) was considered statistically
significant.

**Results**

**Patient Characteristics**

In our previous study, 24 patients undergoing an
AVM resection were assigned a pretreatment compact-
ness score. Eleven patients were male (45.8%). The med-
ian age at the time of AVM resection was 10.5 years
(interquartile range 8–14 years).

**Prior Compactness Score**

A compactness score of 1 was assigned to 6 patients
(25%), a score of 2 to 16 patients (66.7%), and a score of 3
to 2 patients (8.3%). All 4 patients with a recurrent AVM
had a compactness score of 1, and a lower compactness
score was associated with recurrence after resection (\(p =
0.0003\)).\(^1\)\(^,\)\(^1\)\(^1\)

**Intrarater and Interrater Reliability**

The intrarater reliability for the neuroradiologist was

F. A. Frisoli et al.
AVM compactness score reliability

1.0; there was perfect agreement between the initial and repeat scoring. Of the 72 assigned scores from the new study raters (3 raters grading 24 patients), 64 classifications (88.9%) were the same as the gold-standard scores assigned by the attending pediatric neuroradiologist. The κ statistic among the 4 raters was 0.69 (95% CI 0.44–0.89). The interrater reliability between the neuroradiologist and the neuroradiology fellow was 0.59 (95% CI 0.20–0.89). The fellow misclassified two true scores of 1 as scores of 2 and three true scores of 2 as scores of 1. The interrater reliability between the neuroradiologist and the pediatric neurosurgeon was 0.74 (95% CI 0.41–1.0). The surgeon misclassified two true scores of 1 as scores of 2 and one true score of 2 as a score of 3. The 8 discordant scores occurred in 7 patients. There was perfect agreement between the neuroradiologist and the pediatric interventional radiologist (κ = 1.0).

Discussion

The true incidence of recurrence after resection of AVMs in the pediatric population is not well reported in the literature. A few studies suggest that recurrence in children is higher than in adults.\textsuperscript{1,7} Arteriovenous malformations are more prone to hemorrhage in children and thus carry significant morbidity and mortality.\textsuperscript{2,4} There-

<table>
<thead>
<tr>
<th>Compactness Score</th>
<th>Definition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>diffuse*</td>
<td>These are the least compact &amp; are rather diffuse AVMs. There is considerable intervening brain tissue among the anomalous vessels. Some may appear to border on cerebral proliferative angiopathy.</td>
</tr>
<tr>
<td>2</td>
<td>intermediate</td>
<td>The nidus is apparent &amp; can be of variable size, but there is some dispersion w/ many perinidal anomalous vessels.</td>
</tr>
<tr>
<td>3</td>
<td>compact</td>
<td>These are very compact AVMs. They have a prominent nidus w/ minimal surrounding perinidal anomalous vessels.</td>
</tr>
</tbody>
</table>

* A diffuse AVM (compactness score of 1) does not refer to size. These AVMs can be large or small, and the rating is based on the overall compactness of the nidus and perinidal anomalous vessels.

\textbf{Fig. 1.} Examples of AVM compactness Score 1. \textbf{A:} Lateral cerebral angiogram demonstrating a large diffuse AVM (arrows). \textbf{B:} Lateral cerebral angiogram demonstrating a small but diffuse AVM near the vertex (arrows). The AVM does not have a compact nidus. An early draining vein is seen along its superior margin (uppermost arrow). \textbf{C:} Cerebral angiogram showing a large diffuse AVM. There is a more compact portion of the AVM with a large early draining vein, but there is also a larger yet less compact component of abnormal vascularity with large amounts of intervening brain tissue as well (arrows), which make this a diffuse AVM with a compactness score of 1. \textbf{D:} Lateral cerebral angiogram demonstrating another small but diffuse AVM (arrows). A number of narrow-caliber, small, abnormal vessels are seen in the region of the AVM, without a compact nidus.

\textbf{Fig. 2.} Cerebral angiogram demonstrating an AVM with a compactness score of 2.
fore, the ability to predict recurrence after resection has significant implications for monitoring and follow-up imaging. The Spetzler-Martin AVM grading scale has been a long-standing tool used to predict postoperative neurological complications.\(^{13}\) However, this scale was not designed as a tool for predicting recurrence. A study by Du et al.\(^3\) in a cohort of mostly adult patients with AVM found that a noncompact AVM nidus was an independent marker for poor clinical outcomes after resection, but recurrence was not specifically examined in relation to diffuseness. In that study the authors developed a novel computerized method for quantifying the diffuseness of an AVM nidus. Our previous study recently demonstrated that a visual presurgical angiographic compactness score predicts recurrence. In our cohort, all patients who had a recurrence after complete resection had a diffuse AVM (compactness score of 1) graded by a single pediatric neuroradiologist.\(^{13}\) However, the intra- and interrater reliability of the visual compactness score among radiologists and neurosurgeons was not determined.

Accurately identifying diffuse AVMs may prevent the potentially devastating consequences of missing a recurrence by performing adequate follow-up imaging in high-risk patients. However, to be clinically useful, there must be acceptable reliability among practitioners. This study demonstrates that the visual compactness score has moderate to substantial interrater reliability among trained radiologists and neurosurgeons. Both the neuroradiology fellow and neurosurgeon misclassified two cases might improve the interrater reliability, particularly for those with less experience in angiographic interpretation. The interrater reliability was excellent (perfect) between the pediatric neuroradiologist and pediatric interventional radiologist, the 2 raters with the most extensive experience performing and interpreting angiograms. Additionally, the intrarater reliability for the pediatric neuroradiologist was perfect.

### Conclusions

The visual pediatric AVM compactness score was demonstrated to predict recurrent AVM after resection in a recent retrospective cohort.\(^{11}\) In the present study, we demonstrated that the visual compactness score had moderate to substantial interrater reliability among radiologists and neurosurgeons with various levels of training and experience. However, to avoid misclassifications, the visual compactness score would be applied better by those with more extensive angiographic experience, demonstrated by the perfect interrater reliability between the pediatric neuroradiologist and interventional neuroradiologist. In the future, the accuracy and usefulness of the AVM visual compactness score for identifying children with recurrent AVMs should be evaluated in a prospective cohort. The AVM compactness score in children could lead to a reduction in unnecessary angiograms in children with low-risk lesions for recurrence after complete resection.

### Disclosure

Dr. Beslow has been funded by NINDS Grant No. K12-NS049453, NIH Grant No. T32-NS007413, and The L. Morton Morley Funds of The Philadelphia Foundation.

Author contributions to the study and manuscript preparation include the following. Conception and design: Beslow, Lang, Storm. Acquisition of data: Vossough, Cahill, Heuer, Dahnoush, Storm. Analysis and interpretation of data: Frisoli, Lang, Vossough, Beslow. Drafting the article: Frisoli, Lang. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Beslow. Statistical analysis: Beslow. Study supervision: Beslow.

### References

5. Hoh BL, Ogilvy CS, Butler WE, Loeffler JS, Putman CM, Chapman PH: Multimodality treatment of nongalenic arte-
AVM compactness score reliability


Manuscript submitted September 17, 2012. Accepted February 12, 2013. Please include this information when citing this paper: published online March 15, 2013; DOI: 10.3171/2013.2.PEDS12465. Address correspondence to: Lauren A. Beslow, M.D., M.S.C.E., Departments of Pediatrics and Neurology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520, email: lauren.beslow@yale.edu.