Brain arteriovenous malformation in children

Tae Sung Park, M.D.

Department of Neurosurgery, St. Louis Children’s Hospital, Washington University, St. Louis, Missouri

In their retrospective study, Dr. Sanchez-Mejia and co-workers compare outcomes in children and adults who underwent microsurgical resection of brain arteriovenous malformations (AVMs). The study is well intended and offers new information about outcomes in modern surgical treatments of brain AVMs in children. The authors state the following conclusions:

This analysis confirms the observation that children fare better than adults after microsurgical AVM resection. This discrepancy cannot be explained by differences in AVM anatomy, lesion rupture rates, presenting neurological condition, or treatment techniques, leading the authors to infer that neural plasticity may augment surgical tolerance and recovery in children.

Although the conclusions are certainly attractive, a careful review of the study data reveals a number of flaws in the outcome and statistical analyses. Readers should be cautious about agreeing with the main conclusions for the reasons explained forthwith.

First, it is problematic that the authors used the modified Rankin Scale (mRS) to gauge outcomes in children. The Rankin Scale was originally formulated in 1957 to measure handicaps in patients younger than 60 years of age who had experienced strokes. A modified version of the scale, known as the mRS, was developed later, and it has been used widely to measure handicaps in adults who have suffered from stroke. There has been no version of the mRS applicable to children; however in this study, Sanchez-Mejia and colleagues used for the first time a version that has never been validated. Moreover, it is clear that the mRS used for adults and for children is dissimilar. The mRS for the latter includes the following grades and descriptions: neurological deficit that interferes with daily activities but preserves some functional use of the involved limb (Grade 2); no functional use of a limb but able to resume baseline daily activities such as crawling, sitting up, and walking (Grade 3); and no functional use of a limb and cannot resume baseline daily activities (Grade 4). In comparison, the mRS used in an adult cohort includes the following: unable to engage in all previous activities but able to look after own affairs without assistance (Grade 2, slight disability); requiring some help but able to walk without assistance (Grade 3, moderate disability); and unable to walk or attend to own bodily needs without assistance (Grade 4, moderately severe disability). The mRS used in regard to children in the study by Sanchez-Mejia and colleagues focuses on motor weakness rather than the overall handicaps in these patients. In effect, the authors have used two different assessment scales to compare outcomes in the condition of adults and children. A number of stroke assessment scales other than the mRS (for example, the Barthel Index) are available, and the authors are not justified in claiming that their adaptation of the mRS is the most appropriate measure for the assessment of outcomes in regard to children.

Second, a number of questions arise about the statistical analysis in the paper. The authors attempted to determine whether neurological outcome can be explained by age group, demographics, and other clinical characteristics. This question is addressed partially by the data from the multivariate models that appear in Table 5. A comparison of all the demographic and clinical characteristics grouped by neurological outcome is not presented. Statements throughout the paper lead one to infer that these comparisons were performed, but the results are not reported. For example, the authors state that chi-square tests were performed between predictors and outcome variables (that is, mRS scores); however, they report chi-square tests (or univariate comparisons) by age group only, and not by surgical-outcome group. This presentation indicates that univariate logistic regression analysis was used, followed by a multivariate model. The authors also state that the superior outcomes reported for children could not be explained by demographic factors, clinical features, or treatment methods. An analysis of the association of all of these variables with surgical outcome is not reported. Also, because the multivariate models include only a subset of all possible predictors, were the unreported univariate models used to determine which variables to include in the multivariate models? How was the number of possible predictors reduced from all of the predictors listed in Table 2 to the five included in the multivariate model?

The mRS scores at presentation, at final follow up, and at the change between time points were analyzed by three statistical methods: 1) as categorical measures with seven levels, reported in Table 2; 2) as continuous measures, shown
in Table 3; and 3) as collapsed categorical measures with two levels, seen in Table 4. The mRS scores range in value from 0 to 6 where 0 represents a normal neurological condition and increasing values represent increasing neurological deficit. Because mRS scores are inherently rank ordered, they should be analyzed as such. The chi-square analysis in the first method does not account for the natural ordering of values, and the small number of patients in the seven categories makes the chi-square analysis invalid. Because mRS scores appear to be rank ordered rather than continuous, a nonparametric alternative for the second method is more appropriate. Although it is reassuring that conclusions are consistent across analysis methods, the authors should clearly specify that the analysis of the adapted mRS scores is exploratory; thus, mRS scores are analyzed both as ordered variables (second method) and collapsed into meaningful categories (third method). Method 1 should be removed.

Chi-square tests are appropriate to use in determining whether distributions of unordered, categorical variables differ from one another. Some variables in Table 2 are seemingly continuous or rank ordered (for example, AVM measurement and Spetzler–Martin grade). More appropriate statistical tests include the unpaired t-test (for continuous variables) or the Wilcoxon test (for rank-ordered variables). Because AVM measurement is inherently continuous, is there a reason to collapse AVM measurements into three categories, thereby losing precision? Additionally, many cell sample sizes in the cross tabulations are small and have expected counts of less than five, thus rendering chi-square tests invalid. Were exact methods used in cases in which chi-square tests invalid?

I also question how the mean “difference” values in Table 3 were calculated. The adults and children are not paired; thus, it is not possible to calculate the mean between-group difference. Additionally, with five predictors and only three children who experienced a poor final outcome as well as only two whose condition deteriorated, the appropriateness of the multivariate model demonstrated in Table 5 is questionable. Were model fit statistics examined? Caution must be used to avoid over-interpreting results that are based on data relating to three or fewer children who experienced poor outcomes.

The authors’ statement that “this statistical analysis confirmed that large size and location of the AVM in eloquent cortex predicted poor outcomes in children” misinterprets the results tabulated in Table 5. The significance of size and location in eloquent cortex is not specific to children. The appropriate interpretation is the following: “after adjusting for all other variables in the model, older age, larger size, and AVM eloquence are significant predictors of poor final outcome.”

The follow-up period for the children and adults was 5.2 and 2.8 years, respectively. Is the follow-up duration significantly different between groups? If so, it should be included as an additional covariate in the multivariate model to verify that between-group differences in outcomes are unrelated to between-group differences in follow up.

References

RESPONSE: We appreciate the editorial comments of Dr. Park and would like to address his concerns in detail. We hope to clarify our methods so that his comments do not detract from an important message in our paper.

First, concerns are raised about applying the mRS to children. In the UCSF Brain Arteriovenous Malformation Study Project, we adopted it as our principal outcome measure because it provides more gradation toward the higher-functioning portion of the scale than other measures (such as the Glasgow Outcome Scale) and because it provides more sensitive and meaningful information about the impact of surgical therapy. The mRS specifically assesses levels of independence as compared with previous activity levels, which is problematic with children because some authors consider them inherently dependent. Therefore, some adaptation of the mRS is necessary to ensure a uniform system of grading neurological outcomes. We adapted the mRS to children by assessing the extent of neurological deficit and its impact on baseline, age-adjusted activity, rather than focusing on the extent of dependency, a factor more relevant to adult patients after stroke (which the scale was originally designed by Rankin to assess). Our adaptation is minor—it focuses on motor function (the most relevant and measurable function in small children), it was developed by pediatric neurovascular neurologists and pediatric neurosurgeons who had experience in assessing neurological outcomes, and it was uniformly applied by a dedicated research nurse who assesses all outcomes in patients with AVMs. We think that our work exemplifies a single-outcome assessment that is independent of patient age.

Dr. Park suggested using the Barthel Index, which enables outcomes to be assessed by a scoring system calibrated to various activities of daily living such as bowel and bladder continence, grooming, toilet use, feeding, bathing, and dressing. The use of this outcome scale in children would be more problematic than the mRS, because with this index a healthy child whose surgical result was deemed perfect would be assessed as dependent. Unfortunately, there is no uniform, validated system to use in assessing neurological outcomes in patients of all ages. It was our intention to devise a solution, however, and to proceed with addressing an important clinical issue; namely, how age influences outcomes after AVM resection. Our efforts in this paper represent the best available published data on this issue. The editorial comments suggest that clinical neurosurgical research like ours should not be conducted until neurological outcome measures applicable to all ages are developed and validated. We counter by saying that excluding children from such research until these tools are validated is simply not justifiable.

Second, concerns are raised about our statistical methods. Although we addressed many of these issues carefully during the peer-review process, we will enumerate them again because they have been challenged by Dr. Park. 1) Dr. Park states that chi-square tests are invalid if any of the cells or subgroups has an expected count less than five. In those instances, we used Fisher exact tests. 2) The mRS is an ordered categorical variable, and therefore the Wilcoxon rank-sum test could have been used in our analysis. The