

Are your MRI contrast agents cost-effective?

Learn more about generic Gadolinium-Based Contrast Agents.



AJNR

**Further Examination of Diagnostic
Performance in the Context of a Fellows'
Journal Club Article**

R.E. Carter and V.T. Lehman

AJNR Am J Neuroradiol 2012, 33 (6) E96-E97

doi: <https://doi.org/10.3174/ajnr.A3187>

<http://www.ajnr.org/content/33/6/E96>

This information is current as
of April 17, 2024.

Further Examination of Diagnostic Performance in the Context of a Fellows' Journal Club Article

We reviewed the article "The Predictive Value of 3D Time-of-Flight MR Angiography in Assessment of Brain Arteriovenous Malformation Obliteration after Radiosurgery" by Buis et al¹ at a recent neuro-radiology journal club meeting. In this context, the article was thoroughly reviewed with each element carefully critiqued. While other aspects of the study were questioned, in particular the specific MR images used for AVM assessment after radiosurgery and the consistency of use of the abbreviation "PO" to designate "probable obliteration" as a measure of degree of confidence, we have focused our attention in this letter on the reported measures of diagnostic performance. This letter was not written to critique the aforementioned article but rather to highlight teaching points that are made possible by the article.

In the article, the authors define "sensitivity" as "the probability of finding obliteration on MRI₂ among those images demonstrating complete obliteration of DSA_{2c}" and "specificity" as "the probability of finding a patent nidus among those whose images demonstrated no obliterations on DSA_{2c}." Based on these definitions, the reference standard is the DSA diagnosis and the index test is the MR imaging findings. The definitions are considered atypical because sensitivity, as defined, represents the condition of absence instead of the more traditional presence of a diseased condition. Nevertheless, these definitions are mathematically sound.

The estimation of sensitivity and specificity requires binary decisions: The reference standard is positive or negative; the index test is positive or negative. In the usual way, sensitivity is $TP/(TP + FN)$, where TP is the number of true-positive cases (positive index test and positive reference standard) and FN is the number of false-negatives (negative index test and positive reference standard). Likewise, specificity is $TN/(TN + FP)$, where TN is the number of true-negative cases (negative index test and negative reference standard) and FP is the number of false-positives (positive index test and negative reference standard).

In the context of the article, the application of these standard definitions is not straightforward because Table 3 in the article does not use binary decisions. The MR imaging findings are presented as a trichotomous variable with Patent, PO (partial or probable obliteration), and DO (definitive obliteration) categories. To form a binary classification, these 3 distinct categories need to be combined into 2 values: absent or present. The methods do not provide this decision rule, but by using the authors' definition for sensitivity, 1 grouping of the MR imaging findings would be to treat DO as synonymous with "obliteration" so that sensitivity for reader 1 would equal 61.5% (48/78). This calculation does not match the results reported in their Table 4. Sensitivity for reader 1 is reported as 52%. One possible explanation for the difference is that data were combined in a different manner. The only other option is combining PO with DO values to represent obliteration. This yields a sensitivity of 80.8% (63/78) for reader 1. This calculation still does not agree with the results presented in their Table 4. Using a similar strategy, one could continue to perform calculations for reader 2 and other measures of diagnostic performance reported in their Table 4 and reach the conclusion that the numbers presented are not supported by the data in their Table 3. Did the authors make calculation mistakes or is there another explanation?

The explanation for the discrepancy is that the authors have calculated the diagnostic performance summaries by using the MR imaging findings as the reference standard and the DSA results as the

index test with a reversed definition of disease present. Accordingly, one is able to reproduce all numbers in their Table 4 if one combines the Patent and PO categories into reference standard positive (ie, nidus present with certainty or probable certainty) and considers DSA patent as test positive. For example, the "sensitivity" and "specificity" by using these amended definitions for reader 1 would be 52.4% (33/63) and 88.9% (48/54), respectively. Thus while the numbers reported in their Table 4 are reproducible, the meaning of the indices has been altered, with the reversal of the disease-positive and -negative classification and the switching of the reference standard and index test.

This raises the second teaching point: How is 52.4% interpreted if it is not "sensitivity" as defined by the authors? The value actually represents a sample estimate of the positive predictive value (PPV) of MR imaging findings (ie, DSA as the reference standard). Specifically, there would be 33 TP cases and 30 FP cases, so that the PPV would be $33/(33 + 30)$ or 52.4%. This estimate is only valid in a simple random sample design that measures both MR imaging and DSA results on all cases. This is to ensure that the disease prevalence is not altered experimentally. Approximately half of the cases were not included in the analysis, so it would be unreasonable to assume that the disease prevalence was not altered. Practically speaking, the study may be "enriched" with reference standard-positive cases because the observed prevalence is reported as 67% (78/117). When this occurs, the PPV should be estimated by using estimates of sensitivity, specificity, and the disease prevalence in the general screening population by using this formula: $PPV = [Sensitivity \times Pr(Disease)]/[Sensitivity \times Pr(Disease) + (1 - Specificity) \times (1 - Pr(Disease))]$, where $Pr(Disease)$ represents the disease prevalence and Sensitivity and Specificity are in their decimal (probability) forms.

A similar formula exists for NPV.² The Table presents PPV and NPV values for various disease-prevalence values. For these calculations, the sensitivity and specificity are estimated as 61.5% (48/78) and 84.6% (33/39) on the basis of the performance of reader 1. The numbers would be different on the basis of the performance of reader 2. The Table illustrates that the disease prevalence has a profound impact on both PPV and NPV.

In summary, the article by Buis et al¹ emphasizes the need for specific reporting of the decision rules for combining multicategory ratings into the dichotomous ratings required for diagnostic performance calculations. Careful attention to the reference standard and its adjudication is required to interpret sensitivity and specificity correctly. Finally, one must be cautious when interpreting PPV and NPV by ensuring that the disease prevalence is representative of the general screening population and has not been altered through the inclusion/

PPV and NPV for time-of-flight MR imaging as a screening test for complete obliteration for various disease-prevalence values^a

Disease Prevalence	PPV	NPV
0.1	30.8%	95.2%
0.2	50.0%	89.8%
0.3	63.2%	83.7%
0.4	72.7%	76.7%
0.5	80.0%	68.8%
0.6	85.7%	59.5%
0.667 ^b	88.9%	52.3%
0.7	90.3%	48.5%

^a Disease prevalence represents the pretest probability of complete obliteration without the time-of-flight MR imaging results. PPV represents the posttest probability of complete obliteration based on a screening test with 61.5% sensitivity and 84.6% specificity.

^b Sample estimate by reader 1 was 67% (78/117) in Buis et al.¹

exclusion criteria of the study or by being aware that the disease prevalence may be affected by artifacts of missing data.

References

1. Buis DR, Bot JC, Barkhof F, et al. **The predictive value of 3D time-of-flight MR angiography in assessment of brain arteriovenous malformation obliteration after radiosurgery.** *AJNR Am J Neuroradiol* 2012;33:232–38
2. Zhou XH, Obuchowski NA, McClish DK. *Statistical Methods in Diagnostic Medicine*. 2nd ed. Hoboken, New Jersey: John Wiley and Sons; 2011

<http://dx.doi.org/10.3174/ajnr.A3187>

R.E. Carter
Department of Health Sciences Research
V.T. Lehman
Department of Radiology
Mayo Clinic
Rochester, Minnesota