



Published in final edited form as:

Acad Radiol. 2010 September ; 17(9): 1083–1089. doi:10.1016/j.acra.2010.04.025.

Validation of a New Reference Standard for the Diagnosis of Vasospasm

Melissa Reichman, MD¹, Rachel Gold², Edward Greenberg, MD¹, Jana Ivanidze³, Elliot Elias⁴, Joseph Comunale, MD¹, Apostolos J. Tsiouris, MD¹, Carl Johnson, MD¹, and Pina C. Sanelli, MD, MPH^{1,5}

¹Department of Radiology, Weill Cornell Medical College/New York Presbyterian Hospital 525 East 68th Street New York, NY 10065

²New York College of Osteopathic Medicine 1 Northern Boulevard, P.O. Box 8000 Old Westbury, NY 11568

³Ludwig Maximilians University of Munich Geschwister-Scholl-Platz 1 Munich, Germany 80539

⁴Boston University School of Medicine 715 Albany Street Boston, Massachusetts, 02118

⁵Department of Public Health, Weill Cornell Medical College/New York –Presbyterian Hospital 525 East 68th Street New York, NY 10065

Abstract

PURPOSE: The purpose of our study is to perform an internal validation of a new reference standard for vasospasm diagnosis in aneurysmal subarachnoid hemorrhage (A-SAH) patients.

METHODS: A retrospective study was performed on A-SAH patients from 1/2002-5/2009. All patients were applied to this new reference standard using a multi-stage hierarchical approach incorporating clinical and imaging criteria. An internal validation method was performed in two phases to compare the new reference standard with digital subtraction angiography (DSA) and to assess accuracy. In Phase I, the diagnostic outcomes from DSA at the primary level were compared with the secondary/tertiary levels in the reference standard. In Phase II, the new reference standard was compared with chart diagnosis. Accuracy test characteristics, agreement rates, kappa values and bias indices were calculated.

RESULTS: In Phase I (n=85), there was 87% agreement rate, 0.674 kappa and 0.12 bias index. However, there was 100% agreement in patients diagnosed with vasospasm by DSA. Sensitivity, specificity, PPV and NPV are 100%, 61%, 83% and 100% respectively. In phase II (n=137), there was 91% agreement rate, 0.824 kappa, and 0.04 bias index. Sensitivity, specificity, PPV and NPV are 88%, 95%, 96%, and 87% respectively.

CONCLUSIONS: Performing validation methods for a new reference standard is an evolving and on-going process because limitations and bias in the reference standard are identified. Based on the results of this internal validation, a modification in the new reference standard is made at the primary level, resulting in improvement in its accuracy and classification of A-SAH patients.

© 2010 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.

Corresponding Author: Melissa Reichman, MD Department of Radiology New York Presbyterian Hospital 525 East 68th New York, NY 10065 Telephone: (212)-774-1891 Street melissareichman@gmail.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Acute subarachnoid hemorrhage as a result of a ruptured aneurysm is a devastating condition, with an incidence of approximately 10 per 100,000 persons annually (1), associated with as great as 67% patient fatality (2). Delayed cerebral vasospasm is a serious complication of aneurysmal subarachnoid hemorrhage (A-SAH), typically developing 4 to 9 days after the hemorrhagic event. Symptomatic vasospasm has been reported in 22% to 40% of patients and is a significant cause of morbidity and mortality in this patient population (3). Poor clinical outcome is associated with vasospasm leading to permanent neurologic deficits, stroke and death.

Today, developing new technology is being studied to improve early and more accurate diagnosis of vasospasm. Identification of these patients with vasospasm is important for initiation of prompt treatment to prevent stroke and death. In addition, accurate classification of patients without vasospasm is also needed to limit the neurologic and systemic adverse effects associated with treatment of vasospasm. Serious complications can occur in patients incorrectly classified as either false-positive or false-negative in this patient population. Thereby, critical assessment of the classification scheme and reference standard for vasospasm in A-SAH patients has become a primary focus of our research.

Currently, digital subtraction angiography (DSA) is considered the gold standard for angiographic vasospasm, which is diagnosed by imaging findings demonstrating narrowing of large or medium sized intracranial vessels (4). This is often associated with diminished perfusion to the territory distal to the involved vessels. However, other gold standards have been used in the literature to determine vasospasm, such as clinical symptoms, transcranial Doppler ultrasound (TCD) and single photon emission computed tomography (SPECT) (5,6,7). Depending on the gold standard used, a definitional shift in the diagnosis of vasospasm occurs and multiple terms are created; such as “symptomatic vasospasm”, “angiographic vasospasm”, “TCD vasospasm” and “hemodynamic vasospasm”. Thereby, the classification of patients with and without vasospasm will vary according to the gold standard used. This variability may affect patient outcomes by altering treatment decisions and medical management.

Even though several methods are used for the diagnosis of vasospasm, a perfect gold standard does not exist. DSA, clinical exam, TCD and SPECT have their limitations in practice. Alternative methods for a reference standard, such as dispute resolution or composite criterion, may be used when an acceptable gold standard does not exist (8). In our prior work, we have developed a new reference standard using a composite criterion incorporating both clinical and imaging techniques for the diagnosis of vasospasm in A-SAH patients (6). Our new reference standard broadens the definition of vasospasm, by including imaging and clinical terms, in an attempt to improve the classification of patients with and without vasospasm. Treatment decisions and clinical consequences may be affected by a new classification scheme. Thereby, it is important for this new reference standard to undergo a vigorous validation process prior to its implementation into practice.

The purpose of this study is to perform an internal validation of our new multi-stage hierarchical reference standard for vasospasm by evaluating its accuracy in correctly classifying patients with and without vasospasm.

Materials and Methods

Study Population

We performed a retrospective study on consecutive patients admitted to our institution, with the diagnosis of A-SAH from January 2002 – May 2009. Inclusion criterion for the study was an admission diagnosis of A-SAH as determined by chart review. Institutional review board approval was obtained.

Study Design

All A-SAH patients were applied to this new multi-stage hierarchical reference standard in a stepwise manner. An advantage of using this reference standard is that no patients were excluded and all patients underwent the same approach. This reference standard has a weighted design with the strongest evidence for vasospasm at the primary level and the weakest evidence at the tertiary (last) level. Figure 1 is a flow diagram illustrating this multistage hierarchical design. The following is a brief description of the application of the reference standard. A detailed discussion of the advantages and limitations of this reference standard is provided by Reichman et al. (6)

Step 1—At the primary level, DSA is used to determine the diagnostic outcome of vasospasm. The angiographic criterion for vasospasm is based on the degree of vessel narrowing as compared to the normal parent vessel diameter. No vasospasm is defined as no evidence of luminal narrowing; mild vasospasm is defined as less than 50% degree of luminal narrowing; moderate vasospasm as 50%-75% degree of luminal narrowing; and severe vasospasm as greater than 75% degree of luminal narrowing.

Step 2—Patients who did not have a DSA performed during hospitalization, proceeded to the secondary level, using both clinical and imaging sequelae to determine a diagnosis of vasospasm. The clinical criterion for a vasospasm diagnosis at this level is presence of a permanent neurologic deficit on clinical examination, distinct from the deficit at baseline produced by the initial A-SAH. The imaging criterion for a vasospasm diagnosis is delayed infarction present on follow up computed tomography (CT) or magnetic resonance imaging (MRI) of the brain. Delayed infarction is defined as a new infarct on CT or MRI, that occurred within the vasospasm period after day 4, which was not present on CT within 3 days after onset. This criteria effectively excludes patients with brain damage from the initial hemorrhagic event and postoperative complications (5). A patient who fit either or both of these criteria was assigned a vasospasm diagnosis at the secondary level. For patients who did not have a DSA and did not fit either of the above criteria, and importantly did not receive treatment for vasospasm, then a no vasospasm diagnosis was assigned.

Step 3—Patients without DSA, who did not manifest clinical or imaging sequelae of vasospasm, and had received treatment for vasospasm, proceeded to the tertiary level in the reference standard. This level was based on evaluating response-to-treatment in patients who received medical therapy, such as medically induced hypertension, hypervolemia and hemodilution (HHH), to augment cerebral perfusion pressure. Patients who demonstrated an improvement in symptoms and/or clinical examination after medical HHH therapy, as determined by medical record review, were considered responders to appropriate treatment and were assigned a vasospasm diagnosis at the tertiary level. Patients who did not improve after treatment and were found to have another etiology for symptoms, such as hydrocephalus, re-hemorrhage from aneurysm, postoperative infarction, infection, etc., were assigned a no vasospasm diagnosis.

Validation Process

For the initial validation of our reference standard, we performed an internal validation process in two phases. Phase I was designed to evaluate the accuracy of the secondary/tertiary levels in the classification of patients by comparing it with the current gold standard, DSA. Only patients who were diagnosed at the primary level with DSA were then applied to the secondary and/or tertiary levels to determine a vasospasm diagnosis, as discussed above. Phase II was performed to evaluate the accuracy of the new multi-stage hierarchical reference standard as it is intended to be used in practice with all A-SAH patients. The gold standard in Phase II of the study was chart diagnosis of vasospasm. Two research assistants performed a retrospective, blinded chart review of the hospital course, including the discharge summary and progress notes. Additional evaluation of the primary, secondary and tertiary levels was performed individually to determine accuracy at each level.

Statistical Analysis

Test characteristics, including sensitivity, specificity, positive and negative predictive values were calculated in Phase I and II of the study. Agreement rates, kappa values, and bias indices were also performed. Kappa, a chance-adjusted measure of agreement between observers, is affected by the presence of bias between observers and by the prevalence. Therefore, the bias index was used as a measure of variance, which accounts for the differences in assessment of the frequency of occurrence of a condition (9).

Results

Patients

A total of 137 patients were identified for inclusion in this study. Importantly, no patients were excluded from the study. There were 85 patients who had DSA performed during their hospital course and were included in Phase I of the study. However, all 137 patients were included in Phase II. Clinical and demographic data are presented in Table 1.

In Phase I (n=85), using DSA at the primary level as the gold standard, vasospasm was diagnosed in 57 (67%) patients and no vasospasm in 28 (33%) patients (Table 2). The secondary/tertiary levels in the reference standard determined a diagnostic outcome in 65 (76%) patients at the secondary level and 20 (24%) at the tertiary level. At the secondary level, 48 (74%) patients were classified with vasospasm and 17 (26%) patients were no vasospasm. At the tertiary level, 20 (100%) patients were classified with vasospasm. The agreement rate between DSA and the secondary/tertiary levels for the diagnostic outcome was 87% (74/85). In the subgroup of patients with vasospasm on DSA (n=57), there was 100% agreement compared with the secondary/tertiary levels, resulting in 100% sensitivity. However, there was only 39% (11/28) agreement for the subgroup without vasospasm on DSA (n=28), resulting in low specificity of 61%. Positive predictive value was 83% and negative predictive value was 100%. The kappa value was 0.674, considered as substantial strength of agreement according to the Landis and Koch rating scale (10). The bias index, used as a measure of variance, was 0.12.

In Phase II (n=137), the multi-stage hierarchical reference standard as designed for implementation in practice was compared to chart diagnosis. Overall, a diagnostic outcome was determined at the primary level for 85 (62%) patients, secondary level for 45 (33%) patients, and tertiary level for 7 (5%) patients (Table 3). At the primary level, 57 (67%) patients were classified as vasospasm and 28 (33%) patients as no vasospasm. At the secondary level, 6 (13%) patients were classified as vasospasm and 39 (87%) as no vasospasm. At the tertiary level, 7 (100%) patients were classified as vasospasm. According to chart review, vasospasm was determined for 76 (55%) patients and no vasospasm for 61

(45%) (Table 3). The agreement rate between chart review and the new reference standard for a diagnostic outcome was 91% (125/137). Sensitivity, specificity, PPV and NPV are 88%, 95%, 96%, and 87% respectively. The kappa value was 0.824, considered in the “almost perfect” agreement category, according to Landis and Koch (10), with a low bias index of 0.04. In this phase, data analysis was also performed at each individual level in the reference standard and compared with chart diagnosis. At the primary level using DSA, the agreement rate was 88% (75/85) with chart diagnosis. Nine out of the 10 patients in disagreement, were considered as vasospasm by chart diagnosis and were assigned a no vasospasm diagnosis using DSA. At the secondary level using clinical and imaging criteria, the agreement rate was 96% (43/45). The two patients in disagreement were considered as no vasospasm by chart diagnosis, however, both had infarcts seen on follow-up CT that met criteria for a delayed infarction for a vasospasm diagnosis. The tertiary level using response-to-treatment analysis had 100% (7/7) agreement with chart review for a vasospasm diagnosis.

Discussion

Often times, a perfect gold standard does not exist in clinical or research practice. The accuracy of a new diagnostic test determined by using an imperfect gold standard introduces biases and inconsistencies in the results. In clinical practice, gold standards are rarer than one might think (11). For example, colposcopy guided biopsy of the cervix has been considered the gold standard for disease detection of cervical neoplasia for decades, however the sensitivity is only 60%, leading to misclassification of patients with cervical neoplasia as negative for the disease. Using such an imperfect gold standard will bias estimates of accuracy of a new and potentially better diagnostic test by categorizing the additional cases of disease detected as “false positive” test results (8). In this situation, the accuracy of the new diagnostic test is limited by the current gold standard used. Thereby, it is often very difficult to validate a new reference standard and few are actually validated that we commonly use today.

In this study, we have performed an internal validation of a new multistage hierarchical reference standard for the diagnosis of vasospasm, to improve accuracy and classification of A-SAH patients. Internal and external validation methods are important for the complete assessment of a reference standard. Internal validation refers to procedures restricted to a single data set and assesses accuracy in the classification of patients with and without disease. Whereas, external validation methods consider the generalizability of the reference standard to other target populations by evaluating its reproducibility and re-test reliability. A robust reference standard has achieved both validation components with high accuracy in the classification of patients and applicability to other target populations (12).

The results from our study are an initial validation for a novel multistage hierarchical reference standard for the diagnosis of vasospasm in A-SAH patients, incorporating both imaging and clinical criteria. The reference standard includes a weighted design with the strongest evidence of diagnosis at the primary level and the weakest evidence at the tertiary level. Our results support the hierarchical design of the reference standard, where the majority of patients (n=85) are determined for vasospasm diagnosis at the primary level, followed by the secondary level (n=45), and lastly the tertiary level (n=7). In Phase I, primary level patients with DSA were applied to the secondary/tertiary level, and again the hierarchy remains with most patients determined at the secondary level, followed by fewer at the tertiary level.

In Phase I, there was a moderately high agreement rate of 87% for the diagnostic outcomes between DSA and the secondary/tertiary levels in the reference standard. The kappa value of

0.674 is considered as substantial strength of agreement. On further analysis, there was 100% agreement in the subgroup of patients with DSA positive vasospasm. There were 11 discrepant cases with DSA negative for vasospasm, but were assigned a vasospasm diagnosis at the secondary/tertiary levels according to clinical and imaging criteria. In this subgroup, 82% (9/11) patients had a vasospasm diagnosis confirmed on chart review. These discordant results may partly be due to the limited ability of DSA in detecting distal vasospasm. Vasospasm of small, intraparenchymal arteries has been postulated as the cause of delayed neurological deterioration in the absence of large artery vasospasm (5). Review of the literature also reveals that using DSA alone has a low sensitivity of 80% (3) for the detection of vasospasm. Even the combined use of TCD and DSA together predicts occurrence of delayed cerebral infarction with suboptimal sensitivity of 72% and specificity of 68% (13). Lastly, symptomatic vasospasm, which is defined by a neurologic deficit on clinical exam or delayed cerebral ischemia, and not angiographic vasospasm, may partly attribute to these discrepancies observed between DSA and the chart diagnosis (5).

In Phase II, there was a high agreement rate of 91% between our multistage hierarchical reference standard and chart diagnosis. The kappa value of 0.824 is considered in the “almost perfect” agreement category. Importantly, the agreement rate was improved when the secondary and tertiary levels are included in the reference standard. Evaluation of each level individually revealed that the primary level did not actually have the highest agreement (88%) with chart diagnosis because of the subgroup of patients with DSA negative exams, as described above. The secondary and tertiary levels had excellent agreement with chart diagnosis (96% and 100% respectively) supporting the importance of including clinical and imaging criteria in the reference standard design. The tertiary level shows promising results, however, the sample size is too small to make conclusive statements.

The strengths of this new reference standard are that it includes all patients in the A-SAH population, including those with and without symptoms. To our knowledge, it is the first reference standard for vasospasm that incorporates both clinical and imaging criteria. Since vasospasm is a complex disease and its pathophysiology is poorly understood, it has been defined according to both clinical and imaging criteria. Another unique feature of this reference standard is at the tertiary level which incorporates treatment analysis. Considering the effects of treatment is rarely included in a reference standard because most patients who have been treated for the disease would be excluded from the analysis. However, in this patient population prophylactic treatment measures may be used and we developed a reference standard that can be applied to all patients in the A-SAH population.

Performing a validation process for a new reference standard is necessary to determine its accuracy in the classification of patients with and without disease. However, we have also realized its additional value in identifying the limitations and bias that may exist in a reference standard. Thereby, modification of the reference standard can be made to adjust for these limitations and allow for improvement. The results of this study support a revision of the reference standard at the primary level for the subgroup of patients that have DSA negative exams. If the DSA is negative for vasospasm, then the patient is not assigned a no vasospasm diagnosis at the primary level, and instead proceeds to the secondary level (Figure 2). The rationale is based on our data indicating that the some of these patients do indeed have vasospasm based on clinical criteria at the secondary level. A secondary analysis was performed using the revised reference standard to assess its improved accuracy in the classification of A-SAH patients. The 28 patients with DSA negative exams at the primary level were evaluated at the secondary/tertiary levels. The overall agreement rate between the revised reference standard with the chart diagnosis improved to 98.5%.

The major limitations of this study are its retrospective study design relying on chart review for data collection. There was variability in the chart documentation of vasospasm between the discharge summary and the progress notes. Another limitation of the study is the small sample size at the tertiary level. In Phase I, 20 patients were classified at the tertiary level and in Phase II, only 7 patients. A larger prospective study is needed to further support our attempts in validating this level separately. Lastly, this study tests diagnostic accuracy of the reference standard for vasospasm diagnosis, and does not address precision, the second component of validation. An external validation process assessing precision and reproducibility of the reference standard will be evaluated in our future work.

Validation of a new reference standard is an evolving and on-going process. Determining the accuracy of our new reference standard and its ability to correctly classify patients with and without vasospasm in the A-SAH population is the initial step in the validation process. Through internal validation, the limitations and bias in the reference standard are identified. Thus, a modification in the new reference standard was made at the primary level addressing the limitation of using DSA alone in the diagnosis of vasospasm. Secondary analysis revealed improved accuracy and classification of patients using the revised reference standard. Once adequate accuracy has been achieved with the new reference standard, then an external validation process is the next step to determine its generalizability to other target populations by assessing its precision and reproducibility. Importantly, evaluation of the clinical effectiveness of the new reference standard in practice and its impact on treatment decisions and patient outcomes is recommended in the validation process.

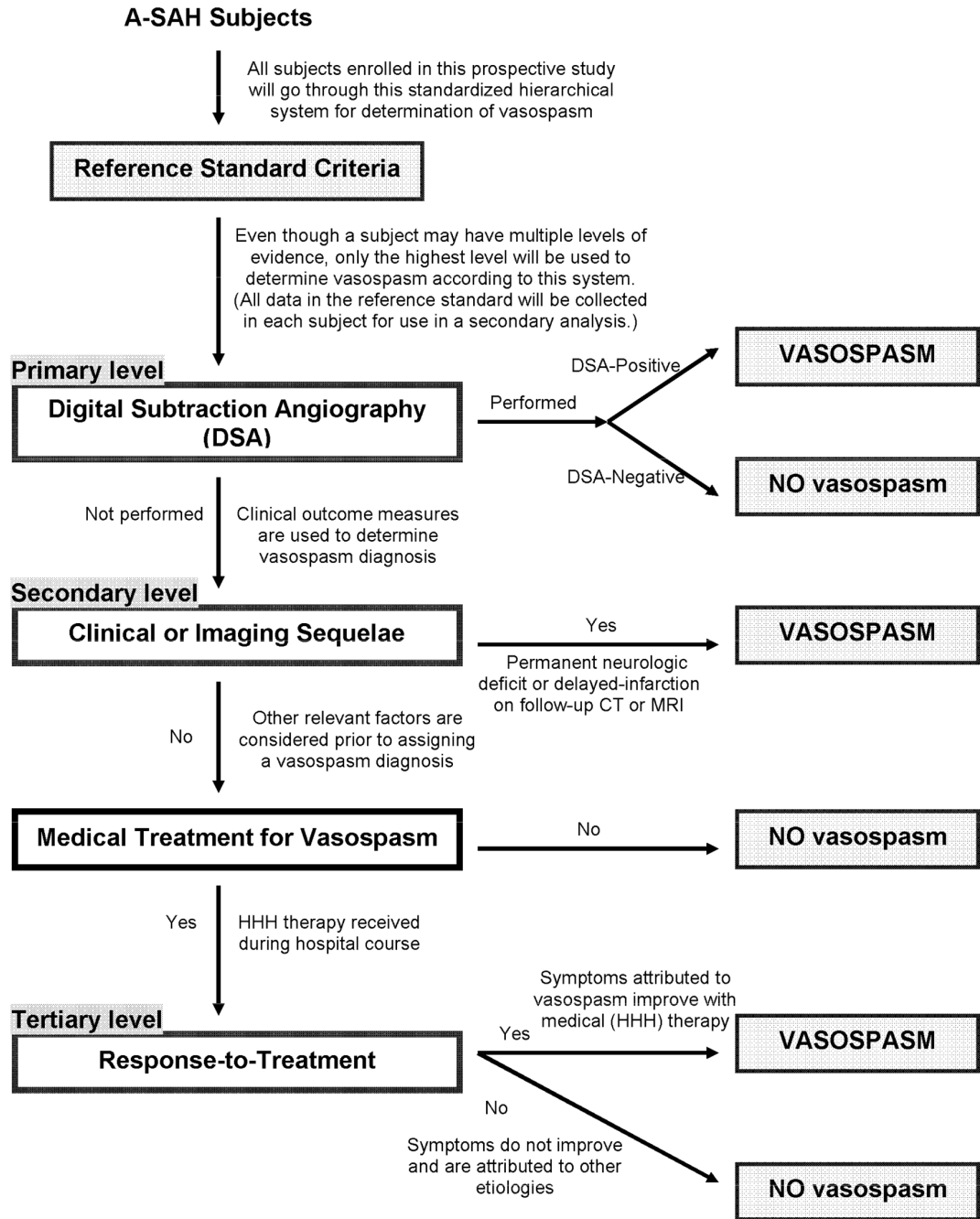
Acknowledgments

This publication was made possible by Grant Number 5K23NS058387-02 from the National Institute of Neurological Disorders and Stroke (NINDS), a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NINDS or NIH.

References

1. Ingal, ITJ.; Whisnant, JP. Epidemiology of subarachnoid hemorrhage. In: Yanagihara, T.; Pepegas, DC.; Atkinson, JLD., editors. Subarachnoid hemorrhage: medical and surgical management. Marcel Dekker; New York, NY: 1998. p. 194-206.
2. Hop JW, Rinkel GJ, Algra A, et al. Case fatality rates and functional outcome after subarachnoid hemorrhage: a systematic review. *Stroke* 1997;28:660–664. [PubMed: 9056628]
3. Suarez J, Qureshi A, Abutaher Y, et al. Symptomatic vasospasm diagnosis after subarachnoid hemorrhage: Evaluation of transcranial Doppler ultrasound and cerebral angiography as related to compromised vascular distribution. *Neurologic Critical Care* 2002:1348–1355.
4. Janardhan, Vallabh; Biondi, A.; Riina, H., et al. Vasospasm in Aneurysmal Subarachnoid Hemorrhage: Diagnosis, Prevention, and Management. *Neuroimaging Clinics of North America* 2006;16:483–496. [PubMed: 16935712]
5. Macdonald R. Management of cerebral edema. *Neurosurg Review* 1998;29:179–193.
6. Reichman M, Greenberg E, Gold R, et al. Developing Patient-centered Outcome Measures for Evaluating Vasospasm in Aneurysmal Subarachnoid Hemorrhage. *Academic Radiology* 2009;16:541–545. [PubMed: 19345894]
7. Zubkov A, Rabinstein A. Medical management of cerebral vasospasm: present and future. *Neurological Research* 2009;31:626–631. [PubMed: 19055879]
8. Alonzo T, Pepe M. Assessing the Accuracy of a New Diagnostic Test When a Gold Standard Does Not Exist. *UW Biostatistics Working Paper Series* 1998:3–32.
9. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *Journal of Clinical Epidemiology* 1993;46:423–429. [PubMed: 8501467]
10. Landis R, Koch G. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174. [PubMed: 843571]

11. Pfeiffer R, Castle P. With or Without a Gold Standard. *Epidemiology* 2005;595–597. [PubMed: 16135933]
12. Altman DG, Royston P. What do we mean by validating a prognostic model? *Statistics in Medicine* 2000;19:453–473. [PubMed: 10694730]
13. Rabinstein A, Friedman J, Weigand S, et al. predictors of cerebral infarction in aneurysmal SAH. *Stroke* 2004;35(8):1862–1866. [PubMed: 15218156]



1

Figure 1.
Flow Chart for Reference Standard To Determine the Study Outcome of Vasospasm Diagnosis
Reprinted from Academic Radiology, 16/5, Reichman, et al., “Developing Patient-centered Outcome Measures for Evaluating Vasospasm in Aneurysmal Subarachnoid Hemorrhage,” pg. 541-545, 2009, with permission from Elsevier

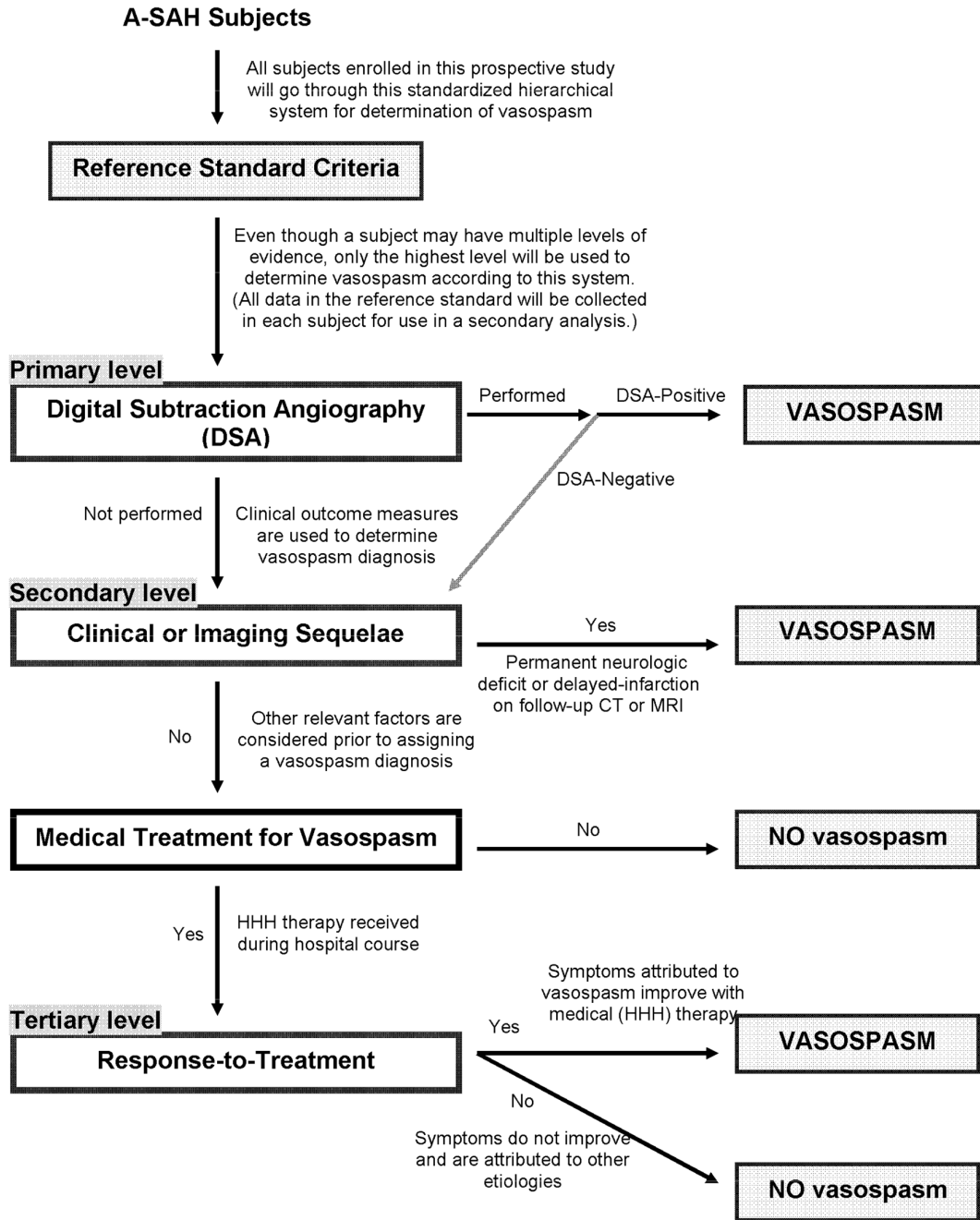


Figure 2. Revised Flow Chart for Reference Standard To Determine the Study Outcome of Vasospasm Diagnosis

Table 1

Demographic Data

	All (n=137)	Vasospasm (n=70)	No Vasospasm (n=67)
Age (years)			
Median	51.5	49	52
Range	24-88	28-88	24-83
Gender			
Male	38 (28%)	21 (30%)	17 (25%)
Female	99 (72%)	49 (70%)	50 (75%)
Aneurysm Location			
Anterior	94 (69%)	52 (74%)	42 (63 %)
Posterior	43 (31%)	18 (26%)	25 (37%)
Treatment Type			
Surgical Clipping	75 (55%)	37 (53%)	37 (55%)
Coil Embolization	59 (43%)	30 (43 %)	29 (43%)
Untreated	3 (2%)	3 (4%)	1 (1%)
Hunt Hess Grade			
Low Grade (1-2)	69(50%)	28 (40%)	42 (59%)
High Grade (3,4,5)	68 (50 %)	42 (60%)	25 (37%)

Table 2

Comparison of Secondary/Tertiary Levels with Primary Level, DSA as Reference Standard.

	Primary Level (DSA)			
		Vasospasm	No Vasospasm	Total
Secondary/Tertiary Levels	Vasospasm	57	11	68
	No Vasospasm	0	17	17
	Total	57	28	85

Table 3

Comparison of Primary/Secondary/Tertiary Levels with Chart Review and Discharge Summary, as reference standard.

	Primary Level (DSA)			Total
		Vasospasm	No Vasospasm	
Multistage Reference Standard	Vasospasm	67	3	70
	No Vasospasm	9	58	67
	Total	76	61	137