Multicenter Accuracy and Interobserver Agreement of Spot Sign Identification in Acute Intracerebral Hemorrhage

Thien J. Huynh, MD; Matthew L. Flaherty, MD; David J. Gladstone, MD, PhD; Joseph P. Broderick, MD; Andrew M. Demchuk, MD; Dar Dowlatshahi, MD, PhD; Atte Meretoja, MD; Stephen M. Davis, MD; Peter J. Mitchell, MBBS; George A. Tomlinson, PhD; Jordan Chenkin, MD; Tze L. Chia, MD; Sean P. Symons, MD, MPH; Richard I. Aviv, MBChB

Background and Purpose—Rapid, accurate, and reliable identification of the computed tomography angiography spot sign is required to identify patients with intracerebral hemorrhage for trials of acute hemostatic therapy. We sought to assess the accuracy and interobserver agreement for spot sign identification.

Methods—A total of 131 neurology, emergency medicine, and neuroradiology staff and fellows underwent imaging certification for spot sign identification before enrolling patients in 3 trials targeting spot-positive intracerebral hemorrhage for hemostatic intervention (STOP-IT, SPOTLIGHT, STOP-AUST). Ten intracerebral hemorrhage cases (spot-positive/negative ratio, 1:1) were presented for evaluation of spot sign presence, number, and mimics. True spot positivity was determined by consensus of 2 experienced neuroradiologists. Diagnostic performance, agreement, and differences by training level were analyzed.

Results—Mean accuracy, sensitivity, and specificity for spot sign identification were 87%, 78%, and 96%, respectively. Overall sensitivity was lower than specificity (P<0.001) because of true spot signs incorrectly perceived as spot mimics. Interobserver agreement for spot sign presence was moderate (κ=0.60). When true spots were correctly identified, 81% correctly identified the presence of single or multiple spots. Median time needed to evaluate the presence of a spot sign was 1.9 minutes (interquartile range, 1.2–3.1 minutes). Diagnostic performance, interobserver agreement, and time needed for spot sign evaluation were similar among staff physicians and fellows.

Conclusions—Accuracy for spot identification is high with opportunity for improvement in spot interpretation sensitivity and interobserver agreement particularly through greater reliance on computed tomography angiography source data and awareness of limitations of multiplanar images. Further prospective study is needed. (Stroke. 2014;45:107-112.)

Key Words: angiography ■ cerebral hemorrhage ■ diagnosis ■ multidetector computed tomography ■ stroke

Intrahematoma contrast density identified on first-pass computed tomography angiography (CTA), coined the spot sign, is associated with increased risk of hematoma expansion, poor functional outcome, and mortality in patients with primary intracerebral hemorrhage (ICH).1-4 The spot sign is currently being studied within 3 multicenter trials to select patients for acute hemostatic therapy. Recombinant factor VIIa is used in the Spot Sign for Predicting and Treating ICH Growth (STOP-IT)5 and Spot Sign Selection of Intracerebral Hemorrhage to Guide Hemosstatic Therapy (SPOTLIGHT)6 studies and tranexamic acid in the Spot Sign and Tranexamic Acid on Preventing ICH Growth—Australasia Trial (STOP-AUST).7 Although the diagnostic performance of the spot sign is previously validated, interobserver agreement and accuracy of spot sign identification and factors associated with their improvement have not been examined in a broad cohort of physicians. To be an effective marker for guiding intervention and patient recruitment into ongoing trials, high spot sign identification accuracy and agreement among acute stroke staff is needed. We sought to evaluate the accuracy and agreement for spot sign identification among readers of various specialties and training levels through use of an online imaging certification module.

Methods

Study Design and Patient Cases

After obtaining local institutional ethics approval, an online imaging module was developed to certify clinical investigators enrolling patients based on spot sign presence in 3 clinical trials: STOP-IT (http://www.stopitstudy.com), SPOTLIGHT (http://www.spotlightstudy.com), and SPOT-AUST (http://www.stopitstudy.com).
mimics shown were cases of ICH secondary to a ruptured microaneurysm. All other cases without spots demonstrated no evidence of spot mimic or cause of secondary ICH. For descriptive purposes, ICH volume was measured using the ABC/2 technique, and spot sign number, maximum density, and axial size were characterized.

Image Acquisition
Noncontrast CT and CTA images were acquired according to our institutional stroke protocol. Noncontrast CT was performed from skull base to the vertex with the following imaging parameters: 120 kVP, 340 mA, collimation 4×5 mm, 1 second per rotation, and table speed of 15 mm per rotation. CTA studies are obtained from the C6 level to the vertex in helical HS mode. CTA parameters were 0.7 ml/kg contrast (maximum of 90 ml through an antecubital vein via a 18–20-gauge angiocatheter), 120 kVP, 270 mA, 1 second per rotation, section thickness 0.625 mm, table speed of 3.75 mm per rotation. CTA contrast bolus timing was obtained using a SmartPrep (GE Healthcare) semiautomated attenuation-triggered technique. Coronal and sagittal maximum intensity projection reformatted images were reconstructed at 7 mm thickness.

Independent Imaging Analysis
Physicians were recruited for imaging module training and certification from July 2008 to May 2012. Specialty and training level (eg, staff physician or fellow) was assessed through a registration form before website use and verified by the study coordinator. After online registration, physicians were provided a brief online teaching module characterizing spot and mimic definitions before the certification module. Certification cases were presented in random order through use of a Flash-based imaging viewer allowing scrolling and windowing for all noncontrast CT, CTA, coronal, and sagittal CTA maximum intensity projection images. Users were not able to measure Hounsfield units directly. Noncontrast CT allowed detection of high density before contrast injection to permit exclusion of nonvascular mimics. For each case, physicians were required to identify spot sign presence or absence, spot sign number, and the presence of spot sign mimics. Time from case presentation to spot sign identification response was recorded using asynchronous javascript and XML. Case review time was not restricted, and readers had the opportunity to review all imaging as many times as needed. Physicians were required to answer ≥80% of spot identification questions correctly to obtain certification for study enrollment. If physicians did not pass the initial certification quiz, they were required to repeat the quiz until a passing grade was obtained in addition to completing supplementary cases. Only physician’s first responses on the certification examination were analyzed for the purpose of this study.

Statistical Analysis
Accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for true spot detection were determined for all physicians and examined by training level. The consensus agreement of 2 experienced neuroradiologists (R.I.A., S.P.S.) was used as reference standard. Differences in diagnostic performance by training level were assessed using the Wilcoxon rank-sum nonparametric test. Interobserver agreement among specialties and training level was determined using the Fleiss multirater k-statistic. Values of k of 0.21 to 0.4, 0.41 to 0.6, 0.61 to 0.8, and 0.81 to 1 were considered fair, moderate, substantial, and nearly perfect, respectively.10 Characteristics of spots with poor interobserver accuracy (≤0.80) and pair agreement (≤0.80 proportion of pairs agreeing) were compared with spots with high accuracy and agreement. Statistical significance was defined as P<0.05 for all tests. All statistical analysis was performed in SAS version 9.2 (SAS Institute, Cary, NC) and R version 3.12.3.

Results
One hundred thirty-one physicians underwent online spot sign imaging training and certification. Physicians were from 28 institutions across the United States (n=13), Canada (n=13), and Australia (n=2). One hundred eighteen (88%) physicians were neurologists (49 staff, 69 fellows), 6 (4.5%) were emergency physicians (3 staff, 3 fellows), and 10 (7.5%) were neuroradiologists (8 staff, 2 fellows). Mean time spent on the educational portion of the imaging module before certification was 17.4±14.2 minutes. Median ICH volume for all certification cases was 32 ml (range, 11–75 ml), and 70% had presence of intraventricular hemorrhage. For the 5 spot-positive cases, median (range) maximum axial spot size was 4 (3.1–10.7) mm, and spot attenuation was 203 (201–496) Hounsfield units. Three of the cases had a single spot, 1 had 3 spots, and 1 had 12 spots.

Physician Diagnostic Performance
Overall mean (95% confidence interval [CI]) for accuracy, sensitivity, and specificity for spot sign identification were 87% (85%–89%), 78% (75%–82%), and 96% (94%–98%), respectively. Overall sensitivity was lower than specificity (P<0.001), and PPV was higher than NPV (96%; [95% CI, 94%–98%] versus 84% [95% CI, 82%–86%; P<0.001]. When true spots were correctly identified, 414 of 514 (81%) responses correctly identified the presence of single or multiple spots. Diagnostic performance and agreement for spot detection are listed in Table 1. Mean (95% CI) PPV and NPV for neurologists was 96% (94%–97%) and 83% (81%–86%).

PPV and NPV for limited sample of emergency physicians and neuroradiologists were 97% (92%–100%) and 85% (75%–95%) and 100% (100%–100%) and 93% (86%–99%), respectively. Accuracy, specificity, and PPV of staff compared with fellows demonstrated no significant difference (P=0.245, 0.983, and 0.916, respectively). There were weak trends toward improved sensitivity (81% [95% CI, 75%–86%] versus 77% [95% CI, 73%–81%]; P=0.123) and NPV (86% [95% CI, 82%–90%] versus 82% [95% CI, 79%–85%]; P=0.114) among staff compared with fellows.

Interobserver Agreement
Agreement for all readers and neurologists was moderate: k=0.60 (95% CI, 0.59–0.61) and k=0.59 (95% CI, 0.57–0.59), respectively. Agreement among staff was similar to fellows...
Limited assessment of neuroradiologists and emergency physicians achieved values of $k=0.88$ and 0.68, respectively.

### Diagnostic Performance, Agreement, and Time to Spot Identification by Case

Accuracy, pair agreement, and time to spot identification by individual examination case are listed in Table 2. Seven cases (2 spot-positive and 5 spot-negative) demonstrated high accuracy ($\geq 80\%$) and pair agreement (Figures 1 and 2), whereas 3 cases (3 spot-positive) were identified as demonstrating low accuracy and agreement (Figures 3 and 4). For the 3 cases with low accuracy, incorrect responses were examined to determine whether readers may have answered the question incorrectly because of the perceived presence of a spot mimic. Of the combined 117 incorrect responses for the 3 questions with low accuracy, 111 (95\%) responses also indicated the presence of a spot mimic. Similarly, for the 2 true spot-positive cases with high accuracy, 22 of 24 (92\%) incorrect responses also indicated the presence of a mimic.

The cases demonstrating the microarteriovenous malformation and aneurysm were correctly identified as spot mimics in 69 of 131 (53\%) and 48 of 131 (41\%) responses, respectively.

Overall median (interquartile range) time to spot identification per case was 1.9 (1.2–3.1) minutes and was similar between staff and fellows ($P=0.673$).

### Discussion

Previous studies of spot sign interobserver agreement have demonstrated substantial to near perfect agreement ranging from $\kappa$-statistic of 0.61 to 0.94$^{1,3,11-13}$; however, little is known about the strength of agreement among a broad range of readers, between physician specialties and training levels, or accuracy compared with a gold standard consensus. Our results demonstrate that there is an overall high accuracy (87\%) for spot identification within a certification context among a large number of readers involved in acute stroke care. We have also identified high specificity (96\%) and PPV (96\%) for physician spot sign interpretation, and that physicians were able to perform spot interpretation rapidly $\approx 2$ to 3 minutes. High specificity and PPV would demonstrate that physicians were highly accurate at correctly identifying the absence of a spot when not truly present and that readers had high accuracy when a spot sign was perceived. These findings are clinically important because high specificity is needed to ensure that patients without spots, and thus at lower risk of hematoma expansion, are not treated with aggressive hemostatic therapy.

### Table 1. Accuracy, Sensitivity, Specificity, and Interobserver Agreement for Spot Sign Presence

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>$\kappa$-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>All physicians</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All (n=131)</td>
<td>87 (85–89)</td>
<td>78 (75–82)</td>
<td>96 (94–98)</td>
<td>0.60 (0.59–0.61)</td>
</tr>
<tr>
<td>Staff (n=57)</td>
<td>88 (85–91)</td>
<td>81 (75–86)</td>
<td>95 (92–99)</td>
<td>0.61 (0.59–0.63)</td>
</tr>
<tr>
<td>Fellow (n=74)</td>
<td>86 (84–89)</td>
<td>77 (73–81)</td>
<td>96 (94–98)</td>
<td>0.60 (0.58–0.61)</td>
</tr>
<tr>
<td>Neurology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All (n=118)</td>
<td>87 (85–89)</td>
<td>78 (74–81)</td>
<td>95 (93–97)</td>
<td>0.58 (0.57–0.59)</td>
</tr>
<tr>
<td>Staff (n=49)</td>
<td>87 (84–90)</td>
<td>79 (73–85)</td>
<td>95 (91–99)</td>
<td>0.58 (0.57–0.60)</td>
</tr>
<tr>
<td>Fellow (n=69)</td>
<td>86 (84–89)</td>
<td>77 (72–81)</td>
<td>96 (93–98)</td>
<td>0.58 (0.57–0.59)</td>
</tr>
</tbody>
</table>

Mean accuracy, sensitivity, specificity, and multirater $\kappa$-statistic are reported with 95\% confidence intervals in parentheses.

### Table 2. Accuracy, Agreement, and Time to Spot Identification by Case

<table>
<thead>
<tr>
<th></th>
<th>Response Frequency</th>
<th>Spot-Positive</th>
<th>Spot-Negative</th>
<th>Accuracy</th>
<th>Proportion of Pairs Agreeing</th>
<th>Median Time (Interquartile Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True spot-positive cases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 1</td>
<td>118</td>
<td>13</td>
<td>0.90</td>
<td>0.82</td>
<td>2.2 (1.3–3.3)</td>
<td></td>
</tr>
<tr>
<td>Case 2</td>
<td>105</td>
<td>26</td>
<td>0.80</td>
<td>0.68</td>
<td>1.8 (1.1–2.9)</td>
<td></td>
</tr>
<tr>
<td>Case 3</td>
<td>120</td>
<td>11</td>
<td>0.92</td>
<td>0.84</td>
<td>2.7 (1.8–4.4)</td>
<td></td>
</tr>
<tr>
<td>Case 4</td>
<td>99</td>
<td>32</td>
<td>0.76</td>
<td>0.63</td>
<td>1.8 (1.1–2.8)</td>
<td></td>
</tr>
<tr>
<td>Case 5</td>
<td>72</td>
<td>59</td>
<td>0.55</td>
<td>0.50</td>
<td>2.2 (1.4–3.4)</td>
<td></td>
</tr>
<tr>
<td>True spot-negative cases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 6</td>
<td>8</td>
<td>123</td>
<td>0.93</td>
<td>0.88</td>
<td>2.2 (1.3–3.3)</td>
<td></td>
</tr>
<tr>
<td>Case 7</td>
<td>8</td>
<td>123</td>
<td>0.93</td>
<td>0.88</td>
<td>1.6 (1.0–2.5)</td>
<td></td>
</tr>
<tr>
<td>Case 8</td>
<td>1</td>
<td>130</td>
<td>0.98</td>
<td>0.98</td>
<td>1.5 (1.1–2.3)</td>
<td></td>
</tr>
<tr>
<td>Case 9</td>
<td>6</td>
<td>125</td>
<td>0.96</td>
<td>0.91</td>
<td>1.8 (1.1–2.6)</td>
<td></td>
</tr>
<tr>
<td>Case 10</td>
<td>5</td>
<td>126</td>
<td>0.96</td>
<td>0.93</td>
<td>1.8 (1.1–2.6)</td>
<td></td>
</tr>
</tbody>
</table>
High PPV ensures that all patients identified as having spots by physicians truly have spots and may benefit from hemostatic therapy.

An important finding of this study, however, was the reduced sensitivity of readers compared with specificity. Reduced spot identification and, therefore, sensitivity may result in failure to treat a true spot-positive patient who potentially would benefit from therapy. Low accuracy and agreement among spot-positive cases, leading to lower sensitivity, also resulted in a lower overall κ-statistic (κ=0.60) than previously reported. Analysis of the potential causes for reduced sensitivity demonstrated that a majority of incorrect spot identification responses in the presence of true spots were associated with the incorrect perception of spot sign mimics. In-depth examination of cases with poor sensitivity revealed that spot signs found in these cases were closely associated with adjacent lenticulostriate arteries or cortical vessels, often with only 1 to 2 mm separation between spot and vessel on axial images, as demonstrated in Figures 3 and 4, respectively. Spot-positive cases with high accuracy did not have similar adjacent vessels. A key defining feature of the spot sign definition is the absence of visible connecting vessels from outside the hematoma.13,9,14 The definition is particularly challenging for anterior basal ganglia and peripheral lobar bleeds where vessels may be in close proximity to a hematoma. Careful evaluation of the association between the focal contrast extravasation and the vessel should be made on thin-section axial CTA source images of ≥0.625 mm to determine whether such a visible connection is present. The absence signifies a spot sign. Although review of coronal and sagittal images is helpful and often routinely provided with CTA source images, a potential pitfall of maximum intensity projection images reconstructed at 7 mm, as performed at our institution, is the inability to clearly separate spots from adjacent vessel. This emphasizes the importance of axial CTA source images for spot sign identification and mimic exclusion. The high rate of perceived mimics in the presence of true spots also demonstrates that readers were easily able to detect contrast density within hematomas but that the difficulty experienced related to deciding between a spot sign versus spot mimic.

Further improvement of spot sign interobserver agreement may be possible with multimodal CT imaging, including post-contrast CT and CT perfusion14–17 visualization of extravasation improving identification. Both imaging protocols also provide a critical opportunity for improving prediction of expansion and clinical outcome and thus may represent the ideal method of spot imaging.15–18 Additional prospective evaluation of delayed or dynamic spot sign imaging in comparison to CTA alone is required.

Examination of accuracy and agreement between staff and fellows yielded no significant differences; however, we did observe minor trends toward improved sensitivity and NPV in staff compared with fellows. Limited data were available for neuroradiologists and emergency physicians, because the small sample size of these groups precluded meaningful comparison with the large cohort of neurologists. Online training is available and is shown to improve spot identification and diagnostic certainty and should be encouraged as a continuing quality improvement process.19,20

A limitation of our study is the limited number of cases in the certification examination. Although inclusion of a greater number of cases would have been ideal, this was balanced with the time required to obtain imaging certification for a large number of investigators. Small sample size results in potentially wide variation in accuracy and agreement depending on the selected cases; however, cases were chosen by
experienced neuroradiologists and thought to be representative of acute ICH cases. The number of cases was balanced by the large number of readers of varied institutions, making our results robust for the cases included. Spot sign prevalence was set to 50%, which is at the higher range of previously reported spot prevalence in acute ICH. This was purposefully chosen to maximize reader experience with spot sign appearance. Although spot sign prevalence approaches 50% when using dynamic or delayed CT techniques, prevalence of 22% to 41% using static CTA in patients presenting <6 hours of onset in practice may alter physician diagnostic accuracy.

A relatively small number of emergency medicine physicians and neuroradiologists participated in the certification process, precluding meaningful comparison with neurologists. Larger numbers of each specialty would need to be recruited to meaningfully compare performance. Because physician accuracy and agreement were assessed immediately after training, our results may represent an overestimate of true physician accuracy. Assessment with either a pretest or post-test after 3 to 6 months may have less impressive results.

To be a clinically useful predictor in the acute setting, understanding factors that affect rapid, accurate, and reliable spot sign interpretation is critical. We have demonstrated an overall high accuracy for spot identification with modest variation between specialty and training level. We have also identified opportunities for improving consensus in spot interpretation, particularly with emphasis on improving sensitivity through careful review of axial CTA source images and avoiding over-reliance on multiplanar images that may falsely suggest a spot mimic due to technical considerations. Further study in a prospective clinical cohort, ideally with inclusion of delayed or dynamic imaging, is needed. The STOP-IT study, which enrolls both spot-positive and spot-negative subjects, will provide an ideal opportunity to examine real-time accuracy and agreement of spot sign identification.

Acknowledgments
We gratefully thank all STOP-IT, SPOTLIGHT, and STOP-AUST study investigators for their participation in the respective studies and for completion of imaging certification. We would also like to thank Janice Carrozzella and Stephanie De Masi for their involvement in coordinating physician recruitment and imaging certification for the STOP-IT and SPOTLIGHT studies, respectively.

Sources of Funding
Dr Huynh is supported through a Canadian Institutes of Health Research Master’s Award and Physician Services Incorporated Resident Research Grant. Drs Huynh and Aviv have received funding for this project from a Sunnybrook SEAC ERC Educational and Scholarship Grant and Canadian Stroke Network Summer Student Grant.

Disclosures
Dr Flaherty serves on an advisory board and as a consultant to CSL Behring and is principal investigator of the STOP-IT study. Dr Gladstone is the principal investigator and Drs Aviv, Demchuk, and Flaherty are members of the executive committee of the SPOTLIGHT trial. Dr Davis is a principal investigator for the STOP-AUST study and Drs Meretoja and Mitchell serve in the executive steering committee. Dr Broderick is principal investigator of the SPOTRIAS Center Grant, which receives study medication for the STOP-IT study from Novo Nordisk. The other authors have no conflicts to report.
References


Multicenter Accuracy and Interobserver Agreement of Spot Sign Identification in Acute Intracerebral Hemorrhage
Thien J. Huynh, Matthew L. Flaherty, David J. Gladstone, Joseph P. Broderick, Andrew M. Demchuk, Dar Dowlatchahi, Atte Meretoja, Stephen M. Davis, Peter J. Mitchell, George A. Tomlinson, Jordan Chenkin, Tze L. Chia, Sean P. Symons and Richard I. Aviv

Stroke. 2014;45:107-112; originally published online November 26, 2013;
doi: 10.1161/STROKEAHA.113.002502
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2013 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/45/1/107