ORIGINAL RESEARCH

Neurobehavioral Deficits After Subarachnoid Hemorrhage in Mice: Sensitivity Analysis and Development of a New Composite Score

Kanako Matsumura, BS;* T. Peeyush Kumar, PhD;* Tejesh Guddanti, BS; Yuanqing Yan, PhD; Spiros L. Blackburn, MD; Devin W. McBride, PhD

Background—Because of the failure of numerous clinical trials, various recommendations have been made to improve the usefulness of preclinical studies. Specifically, the STAIR (Stroke Therapy Academic Industry Roundtable) recommendations highlighted functional outcome as a critical measure. Recent reviews of experimental subarachnoid hemorrhage (SAH) studies have brought to light the numerous neurobehavioral scoring systems that are used in preclinical SAH studies. To gain insight into the utility of these scoring systems, as well as to identify a scoring system that best captures the deficits caused by SAH in mice, we designed the current study.

Methods and Results—Adult male C57BL/6J mice were used. One cohort of mice was randomly allocated to either sham or SAH and had functional testing performed on days 1 to 3 post-SAH using the modified Bederson Score, Katz Score, Garcia Neuroscore, and Parra Neuroscore, as well as 21 individual subtests. A new composite neuroscore was developed using the 8 most diagnostically accurate subtests. To validate the use of the developed composite neuroscore, another cohort of mice was randomly assigned to either the sham or SAH group and neurobehavior was evaluated on days 1 to 3, 5, and 7 after injury. Receiver operating characteristic curves were used to analyze the diagnostic accuracy of each scoring system, as well as the subtests. Of the 4 published scoring systems, the Parra Neuroscore was diagnostically accurate for SAH injury in mice versus the modified Bederson and Katz Scores, but not the Garcia Neuroscore. However, the newly developed composite neuroscore was found to be statistically more diagnostically accurate than even the Parra Neuroscore.

Conclusions—The findings of this study promote use of the newly developed composite neuroscore for experimental SAH studies in mice. **(*J Am Heart Assoc*. 2019;8:e011699. DOI: 10.1161/JAHA.118.011699.)**

Key Words: basic science • neurobehavior • neuroscore • outcomes research • subarachnoid hemorrhage

Subarachnoid hemorrhage (SAH) is a devastating stroke subtype that is associated with severe morbidity and mortality rates of ≈25%.[1] With the high number of clinical trials failing to improve patient outcome despite reducing the pathological consequences of SAH, improved experimental studies (the basis of most trials) are called for. Several roundtables have been held to develop guidelines directed at improving preclinical and translational studies. One particular recommendation from these meetings is the proper selection and use of neurobehavioral testing.[2–5] Although neurobehavior testing is more standardized for experimental ischemic stroke studies (ie, it is known which tests are sensitive and specific for ischemic injury), preclinical SAH studies typically do not use functional outcome tests that can be compared between studies. The lack of consistent and robust neurobehavioral tests for experimental SAH studies has been highlighted in 2 recent reviews.[6,7] However, despite the reviews and call for determining the most useful and appropriate functional tests for detecting injury after SAH, the work has not been performed.

To this end, the current study was designed to specifically investigate the utility of several functional scoring systems for evaluating deficits in mice subjected to SAH via endovascular perforation. Herein, the behavioral performance of mice after SAH is assessed using 4 scoring systems that have been applied to SAH studies without proper validation. Because the Parra Neuroscore was developed for SAH, we hypothesized that the Parra Neuroscore would be the most appropriate scoring system for detecting functional deficits in mice after

From The Vivian L. Smith Department of Neurosurgery, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX.

Accompanying Data S1, Figures S1 through S6, and Tables S1 through S16 are available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.118.011699

*Dr Matsumara and Dr Kumar are co–first authors.

Correspondence to: Devin W. McBride, PhD, The Vivian L. Smith Department of Neurosurgery, The University of Texas Health Science Center at Houston, Houston, TX. E-mail: devin.w.mcbride@uth.tmc.edu

Received December 6, 2018; accepted January 15, 2019.

© 2019 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Clinical Perspective

What Is New?

- We examine the utility and perform a comparison analysis of 4 widely used scoring systems for identifying functional deficits in mice after subarachnoid hemorrhage.
- We found that each of the existing scoring systems has subtests that are not diagnostically accurate at identifying functional deficits after subarachnoid hemorrhage.
- We therefore developed a new composite score using 8 subtests, which are the most diagnostically accurate for subarachnoid hemorrhage deficits.

What Are the Clinical Implications?

- Because a major concern with patients is functional outcome, it is imperative to have diagnostically accurate behavioral tests when performing preclinical studies.
- We developed a new composite neuroscore that is significantly more diagnostically accurate for detecting deficits after subarachnoid hemorrhage compared with 4 widely used scoring systems.

SAH. Our second hypothesis was that, using the most diagnostically accurate individual subtests for sensory-motor function, a new composite neuroscore could be developed that would have greater diagnostic accuracy for injury after SAH in mice than any of the other scoring systems.

Materials and Methods

The experiments were approved by the Animal Welfare Committee at the University of Texas Health Science Center at Houston, were conducted in compliance with the *NIH Guidelines for the Use of Animals in Neuroscience Research*, and are reported in compliance with the Animal Research: Reporting in Vivo Experiments guidelines. Data are available on reasonable request.

Study Design

Fifty-five adult male C57BL/6J mice (28–34 g, Jackson Labs) were used in all experiments. Animals were housed in a humidity- and temperature-controlled room with a 12-hour light-dark cycle. Animals were given ad libitum access to food and water. Animals were randomized (electronically generated) into either the sham or the SAH group, according to sample size calculations. The same surgeon (D.W.M.) performed SAH and sham surgeries, and all animals were treated with the same amount of buprenorphine and saline (on the day of surgery). All investigators responsible for functional assessment, measurement of outcomes, and data analysis were blinded to the experimental groups.

The 3-day study

For the 3-day study, animals were randomly assigned into either the sham (n=16) or the SAH (n=21) group before performing any surgical procedures. Sample size estimation was conducted using data from previous publications. For 2 groups, sample size calculations were as follows. An α=0.05 and power=0.80 were assumed for all cases. The sample sizes were estimated to be as follows: n=15 to 20 for the modified Bederson Score[8,9]; n=9 to 20 for day 1 (minimum difference in means=25 and SD=18[10]; also see Bermueller et al[8] and Thal et al[9]) and n=20 to 22 for day 2 (minimum difference in means=13 and SD=15[10]; also see Thal et al[9]) for the Katz Score; n=6 to 24 for the Garcia Neuroscore (minimum difference in means=6 and SD=3.16[11]; also see data of others[12–16]); and n=3 for the Parra Neuroscore (minimum difference in means=16 and SD=5 [approximated][17]).

Overall, sample sizes were estimated to be n=3 to 24, with most studies indicating a sample size of 9 to 20 is needed. Therefore, in this study, we chose to use a sample size of n=16 per group. Because the mortality of mice after endovascular perforation is ≈20% to 25%, we randomly allocated 21 mice into the SAH group (assuming 5 fatalities) and 16 mice into the sham group.

The 7-day study

After the completion of the 3-day study, analysis of the neurobehavioral performances, and development of a new composite neuroscore for SAH, we performed a 7-day study using a new cohort of mice to validate the use of the developed composite neuroscore. For the 7-day study, animals were randomly assigned into either the sham (n=8) or the SAH (n=10) group before performing any surgical procedures. Sample size was calculated using the mean difference (10.11) and SD of the population (0.4632) from the data of the developed neuroscore at 24 hours after SAH; α was set to 0.05 and power was set to 0.80 using the sample calculation for a longitudinal study,[18] with the added 15% for nonparametric tests.[19] This calculation indicated that 8 mice per group were needed to test for statistical significance. Assuming 1 to 2 moralities suggested that the SAH group needed 9 to 10 mice.

SAH Model

SAH was induced in mice, as previously described.[20] Briefly, mice were anesthetized with isoflurane (induction, 5% isoflurane; maintained, 1.5%–2.5% isoflurane) delivered in oxygen (1 L/min). Buprenorphine was injected SC (0.05–0.1 mg/kg). The surgical site was shaved, and bupivacaine (2 mg/kg) was injected SC near the midline in the surgical site. The animal was placed supine, and the surgical site was sterilized with alternating wipes of betadine and 70% ethanol (3 times). A

vertical incision was made along the midline of the neck, and the external carotid and common carotid arteries were isolated using blunt dissection. The external carotid artery was ligated, leaving a stump. Then, the internal carotid artery was isolated. Vessel clips were placed on the common carotid and internal carotid arteries to momentarily stop blood flow. A cut was made in the external carotid artery stump, and a 5 to 0 monofilament nylon suture was inserted through the opening of the external carotid artery stump. The vessel clip was removed from the internal carotid artery, and the suture was advanced through the internal carotid artery until resistance was felt (≈8 mm). The suture was advanced 1 mm further to perforate the vessel, inducing SAH. The common carotid artery clip was removed, and the suture was immediately withdrawn. Then, the external carotid artery stump was ligated closed and the neck incision was sutured. Antibiotic was applied to the skin, and the animal was allowed to recover. Afterwards, mice were placed back into their home cages and housed in groups of 1 to 5 mice per cage. Analgesic and saline were given twice a day for 3 days, as necessary, for mice subjected to SAH. Animals allocated into the sham group underwent all the same surgical procedures and were given buprenorphine, bupivacaine, and saline on the day of surgery. No sham animals received buprenorphine or saline on the days after surgery.

Neurobehavioral Performance

All animals were used to test for sensorimotor function by 3 independent, blinded scorers (K.M., T.P.K., and D.W.M.). All tests were performed in the same order for every animal. All 3 assessors performed the behavior together (with no interassessor interactions) on each mouse to minimize fatigue and potential learning or boredom. All tests were done at the same time (except the beam walking test), in the same order, at the same time of day for each mouse. After all the neuroscore subtests, including Katz and modified Bederson, the mice were given a 3- to 5-minute break before performing the beam walking tasks. Beam walking was performed 1 time, and all beam walking scoring (ie, for Katz and the neuroscore) was recorded at the same time. All mice had neurobehavioral performance assessed by the below tests 1 day before SAH surgery to confirm that all animals were statistically indistinguishable (data not shown).

For the 3-day study, mice underwent daily testing. For the 7-day study, mice had their neurobehavior tested on days 1 to 3, 5, and 7 after SAH. All animals were euthanized after completing the final day of neurobehavior. The following tests were performed.

Coordination and Balance Tests

Spontaneous activity: the mouse is observed for 3 minutes on its ability to reach and explore (rear) all 4 walls of its environment (Table 1) [21]. Forepaw outstretching: the mouse is held by its tail and allowed to walk using its forepaws only (ie, keeping the hind limbs suspended in the air). Climbing: the mouse is placed on an inclined plane with equally spaced rungs to observe its ability to climb for 1 minute. Balance: while the mouse is freely moving, the animal is observed, and any balance deficits (eg, swaying and stumbling) are recorded. Lateral turning: the mouse is suspended in air (head down) and should be able to turn toward both sides to reach for its tail. Walking: the mouse is allowed to freely walk around and is observed for the ability to turn to the left and right. Beam walking: the mouse is placed on each platform (at both ends of a 1.5-cm round rod) for 30 seconds. Then, the mouse is placed perpendicular to the center of the rod and given 1 minute for traversing the beam.

Table 1. Coordination and Balance Tests

	Score			
Subtest	0	1	2	3
Spontaneous activity (3 min)	No movement	Minimal movement	Touches 1–2 walls	Touches 3–4 walls
Forepaw outstretching	No movement	Moves in circles	Moves to one side	Straight/curved path
Climbing (1 min)	Weak grip and fall down	Climbs but does not reach top and weak grip	Climbs to top and weak grip or does not climb to top and good grip	Climbs to top and strong grip
Balance	Tumbles	Stands, sways	Sways as walking	Changes position easily
Lateral turning	No turning	Unequal turning	Bilateral turning, equal but <45°	Bilateral turning, equal and >45°
Walking	No turning	Unequal turning	Bilateral turning <45°	Bilateral turning >45°
Beam walking (1 min)	Hug or fall <10 s	Stand in one spot or move but fall <25 s	Stays on beam and traverses the beam	Walks onto the platform

Scoring criteria are slightly modified from McBride et al.[21] Score is from 0 (maximum deficits) to 3 (no deficits).

Table 2. Posture and Strength Tests

Subtest	Score		
	0	1	2
Ptosis	Severe	Slight	None
Dyspnea	Gasping	Slight	None
Facial weakness	Severe	Slight	None

Scoring criteria are slightly modified from McBride et al.[21] Score is from 0 (maximum deficits) to 2 (no deficits).

Posture and Strength Tests

Ptosis: both eyes of the mouse are observed for drooping (Table 2) [21]. Dyspnea: the mouse is assessed for difficulty breathing. Facial weakness: the mouse is observed for any facial drooping, asymmetry in facial expressions/jaw, and weakened response to cheek touch.

Reflex Tests

Side stroking: a cotton swab is used to stroke each side of the mouse's body to observe a response (head turning, whisker movement, or flight) (Table 3) [21]. Vibrissae touch: a cotton swab is moved from the rear to stimulate the vibrissae and observe a response. Visual: a cotton swab is advanced toward each eye and is observed for any elicited response. Olfactory: a cotton swab is dipped in honey and advanced toward the mouse's nose to look for any sniffing or exploration. Tactile: the wood end of a cotton swab is used to poke the tops of each hind paw. Postural: the mouse is placed in a cage, and the cage is moved rapidly downwards. Sound: observer clapped. Righting: the mouse is placed on a grid and allowed to grip. The grid is moved so that the mouse is facing downwards, and the time it takes for the mouse to right itself (ie, turn upwards) is recorded (repeated 4 times).

Limb Use Tests

Limb extension: the mouse is held in the air by its tail and observed for limb extension (Table 4) [21]. Forelimb use: while the mouse is freely moving, the use of the forelimbs is observed for any deficit (inability to use or stiffness). Hind limb use: while the mouse is freely moving, the use of the hind limbs is observed for any deficit.

In addition to the above individual tests, which comprise various composite neuroscores, we also assessed the animals' performances using the modified Bederson Score, Katz Score, Garcia Neuroscore, and Parra Neurosore. The modified Bederson Score tests forelimb extension and mobility (Table 5).[8,9] The Katz Score combines 13 subtests assessing general deficits, reflexes, and sensorimotor and coordination deficits (Table 6).[9,22] The total score for the Katz Score is 0 (no deficits) to 100 (maximum deficits). The Garcia Neuroscore is composed of 6 subtests (described above): spontaneous activity, limb extension, forepaw outstretching, climbing, side stroking, and vibrissae touch (Table 7).[23] The total score for the Garcia Neuroscore ranges from 0 (maximum deficits) to 18 (no deficits). The Parra Neuroscore is made up of 9 subtests: spontaneous activity, limb extension, climbing, beam walking, visual, olfactory, tactile, side stroking, and vibrissae touch (Table 8).[17] The total score for the Parra Neuroscore ranges from 5 (maximum deficits) to 27 (no deficits).

Statistical Analysis

All tests in this study were 2 sided. All data are presented as individual data points with the median shown. Normality and homoscedasticity were tested for, but not met, for all data within. Single time point data (ie, behavioral data on day 1 after SAH) were analyzed with the Mann-Whitney *U* test. The

Table 3. Reflex Tests

Subtest	Score			
	0	1	2	3
Side stroking	No response	Unilateral response	Bilateral weak response or strong ipsilateral response	Strong bilateral response
Vibrissae touch (stroke whiskers)	No response	Unilateral response	Weak bilateral response	Strong bilateral response
Visual (tip toward each eye)	No response	Unilateral response	Weak bilateral response	Strong/rapid bilateral response
Olfactory	No sniffing	Brief sniff	Sniff >2 s	…
Tactile (poke top of paws)	No response	Delayed withdrawal	Immediate withdrawal	…
Postural reflex (sudden drop)	Absent	…	…	Present
Sound reflex	Absent	Delayed response	Rapid response	…
Righting reflex (4 trials, 15 s)	Absent	Rights <15 s but >10 s	Rights <10 s but >5 s	Rights <5 s

Scoring criteria are slightly modified from McBride et al.[21] Score is from 0 (maximum deficits) to 3 (no deficits).

Table 4. Limb Use Tests

Subtest	Score			
	0	1	2	3
Limb extension	No movement	Minimal movement	Abnormal forelimb walk	Contralateral forelimbs and hind limbs completely extended
Forelimb use	Severe bilateral deficits	Severe unilateral deficits	Slight deficits (unilateral or bilateral)	None
Hind limb use	Severe bilateral deficits	Severe unilateral deficits	Slight deficits (unilateral or bilateral)	None

Scoring criteria are slightly modified from McBride et al.[21] Score is from 0 (maximum deficits) to 3 (no deficits).

longitudinal data (ie, 3- and 7-day studies) were analyzed with a Scheirer-Ray-Hare test (2-way ANOVA on ranks) and a mixed-effect model. For the factor(s) that were found to be significant by these tests, multiple comparisons were made: the injury factor was analyzed using the Mann-Whitney *U* test, and the time factor was analyzed using the Friedman test (1-way ANOVA on rank with repeated measures) (followed by a Bonferroni post hoc correction for multiple comparisons). GraphPad Prism 6 (La Jolla, CA), SigmaPlot 11.0 (SysStat, Germany), MedCalc Statistical Software v18.5 (Ostend, Belgium), and the Real Statistics Resource Pack software (Release 4.3) were all used for analyzing and graphing data.

Receiver operating characteristic curve analysis

Receiver operating characteristic (ROC) curves were created (GraphPad Prism 6) for each individual subtest to identify the diagnostic accuracy each subtest provides for correctly labeling a mouse into its true group (ie, how sensitive and specific is the subtest for labeling a sham mouse as a sham [rather than incorrectly identifying it as an SAH mouse] and vice versa).[24]

In addition, to assess the diagnostic accuracy of the various neurobehavioral scoring systems tested herein, ROC curves were generated for the modified Bederson and Katz Scores and the Garcia and Parra Neuroscores. To test for statistical significance between the ROC curves of the scoring systems, the method of Hanley and McNeil was used to calculate the *z* score[25]:

$$z = \frac{A_1 - A_2}{\sqrt{SE_1^2 + SE_2^2 - 2rSE_1SE_2}}$$

where A_1 and A_2 are the areas under the curves for scoring systems 1 and 2, respectively; SE_1 and SE_2 are the standard errors of the ROC area for scoring systems 1 and 2, respectively; and *r* is the estimated correlation between the 2 scoring systems. To determine *r*, first 2 intermediate correlation coefficients need to be calculated: the correlation between scoring system 1 and scoring system 2 for the sham group and the correlation between scoring system 1 and scoring system 2 for the SAH group. GraphPad Prism was used to compute these 2 Spearman correlation coefficients. Then, using the table provided by Hanley and McNeil, the correlation between the 2 scoring systems (*r*) was estimated.[25] Finally, after the *z* score between the 2 scoring systems was calculated, the *P* value was determined from a *z*-score table.

Sample size estimation to determine the utility of each neuroscoring system

Sample size analysis for each of the neurobehavioral performance scoring systems used within was performed using the means and SDs from the 24-hour functional performance data. The mean difference between the means for the sham and SAH groups was determined. The SD of the population was calculated as follows:

$$SD_{pop} = \sqrt{\frac{\left((n_{sham} - 1)(SD_{sham})^2 + (n_{SAH} - 1)(SD_{SAH})^2\right)}{((n_{sham} - 1)(n_{SAH} - 1))}}$$

where n_{sham} and n_{SAH} are the number of sham and SAH animals, respectively; and SD_{sham} and SD_{SAH} are the SDs of the sham and SAH groups, respectively. For each scoring system (ie, modified Bederson, Katz, Garcia, and Parra), the mean difference and SD of the population were used in SigmaPlot's *t*-test sample size calculator using a desired

Table 5. Modified Bederson Score

Score	Description
5	Mouse held by tail had normal forelimb extension
4	Mouse with consistent flexion of forelimb on either side and adduction and internal rotation of shoulder
3	Mice were allowed to grip paper with forepaws, and gently pushed forward with pressure against the forepaw shoulders; reduced resistance on paretic side was graded 3
2	Forelimb walking: animals circled
1	Spontaneous circling when allowed to walk normally on floor
0	No spontaneous motion

Score is from 0 (maximum deficits) to 5 (no deficits).[8,9]

Table 6. Katz Score

Deficit	Scoring Criteria and Value	Scoring Criteria and Value	Scoring Criteria and Value
General deficit			
Consciousness	Explores spontaneously (0 points)	No attempt (20 points)	. . .
Respirations	Normal (0 points)	Abnormal (20 points)	. . .
Cranial nerve reflexes			
Olfactory	Present (0 points)	Absent (4 points)	. . .
Vision (follows tip)	Present (0 points)	Absent (4 points)	. . .
Corneal reflex	Present (0 points)	Absent (4 points)	. . .
Whisker movement	Present (0 points)	Absent (4 points)	. . .
Sound reflex	Present (0 points)	Absent (4 points)	. . .
Motor deficit			
Leg/tail	Normal (0 points)	Stiff (5 points)	Paralyzed (10 points)
Sensory deficit			
Leg/tail (pinching)	Present (0 points)	Absent (10 points)	. . .
Coordination			
Beam walking	Present (0 points)	Absent (5 points)	. . .
Placing test (forelimb stretching)	Present (0 points)	Absent (5 points)	. . .
Righting reflex	Present (0 points)	Absent (5 points)	. . .
Stopping at table edge	Present (0 points)	Absent (5 points)	. . .

Score range is from 0 (no deficits) to 100 (maximum deficits).[9] Reprinted from Katz et al,[22] with permission. Copyright ©1995, SAGE Publications.

power of 0.8 and an α of 0.05 to estimate the sample size required to test for statistical significance between sham and SAH groups. This process was mirrored using SigmaPlot's ANOVA sample size calculator for group sizes of 3 to 5. Finally, because the sample sizes estimated by SigmaPlot reflect the assumptions of a parametric test, to estimate the sample sizes required to test for statistical significance using a nonparametric test (ie, ANOVA on ranks or Mann-Whitney), each sample size calculated by SigmaPlot was increased by 15%.[19]

Interoperator variability

Interoperator variation was calculated for each of the scoring systems: modified Bederson Score, Katz Score, Garcia

Table 7. Garcia Neuroscore

	Score			
Subtest	0	1	2	3
Spontaneous activity (3 min)	No movement	Minimal movement	Touches 1–2 walls	Touches 3–4 walls while standing on hind limbs
Side stroking	No response	Unilateral response	Bilateral weak response or strong ipsilateral response	Strong bilateral response
Vibrissae touch (stroke whiskers)	No response	Unilateral response	Weak bilateral response	Strong bilateral response
Limb extension	Both contralateral; limb completely flexed	One contralateral; limb extended and other flexed	Mid flexion of either contralateral limb	Contralateral forelimbs and hind limbs completely extended
Forepaw outstretching	No movement	Moves in circles	Moves to one side	Straight/curved path
Climbing (1 min)	Weak grip and fall down	Climbs but does not reach top and weak grip	Climbs to top and weak grip or does not climb to top and good grip	Climbs to top and strong grip

The scoring criteria are slightly modified from the original scoring criteria proposed by Garcia et al.[23] Score range is from 0 (maximum deficits) to 18 (no deficits).

Table 8. Parra Neuroscore

Subtest	Score			
	0	1	2	3
Spontaneous activity (3 min)	No movement	Minimal movement	Touches 1–2 walls	Touches 3–4 walls
Side stroking	…	No response	Unilateral response	Bilateral response
Vibrissae touch (stroke whiskers)	…	No response	Unilateral response	Bilateral response
Visual (tip toward each eye)	…	No response	Unilateral response	Bilateral response
Olfactory	…	No sniffing	Brief sniff	Sniff >2 s
Tactile (poke top of paws)	…	No withdrawal	Delayed withdrawal	Immediate withdrawal
Limb extension	No movement	Minimal movement	Abnormal forelimb walk	Contralateral forelimbs and hind limbs completely extended
Climbing (1 min)	Falls down	Holds <4 s	Holds on but no movement	Climbs
Beam walking	Falls <2 s	Falls >2 s	Holds but no movement	Walks

The scoring criteria are slightly modified from the original scoring criteria proposed by Han et al.[17] Score range is from 5 (maximum deficits) to 27 (no deficits). The side stroking and beam walking subtests in the Parra Neuroscore were originally called proprioception and balance, respectively.

Neuroscore, and Parra Neuroscore. Interoperator variations were computed using a weighted κ statistic (MedCalc).

Developing a new composite neuroscore for SAH

Two methods were used to develop a new composite neuroscore. First, after computing the ROC curves for each subtest, we used threshold for the area under the curve equal to 0.80 for selecting the "best" subtests for a new neuroscore. We also used a variable selection procedure through Lasso regression to identify the most important subtests (based on a frequency >150, Figure S1). The "best/most important" subtests identified by these methods were combined into a new neuroscore (Table 9). This composite neuroscore was then subjected to the rigorous statistical analysis that each of the other scoring systems underwent using the data from the 3-day study (ie, ROC curve analysis, sample size determination, and inter-operator variability). This neuroscore was validated by performing a 7-day study.

Results

Mortality rates were 0% (0/16) for sham and 26% (8/31) for SAH. For the 3-day study, 1 mouse died within the first

Table 9. New Composite Neuroscore for Evaluating Functional Deficits After SAH in Mice

Subtest	Score			
	0	1	2	3
Spontaneous activity (3 min)	No movement	Minimal movement	Touches 1–2 walls	Touches 3–4 walls while standing on hind limbs
Climbing (1 min)	Weak grip and falls down	Climbs but does not reach top and weak grip	Climbs to top and weak grip or does not climb to top and good grip	Climbs to top and strong grip
Balance	Tumbles	Stands, sways	Sways as walking	Changes position easily
Side stroking	No response	Unilateral response	Bilateral weak response or strong ipsilateral response	Strong bilateral response
Vibrissae touch (stroke whiskers)	No response	Unilateral response	Weak bilateral response	Strong bilateral response
Visual (tip toward each eye)	No response	Unilateral response	Weak bilateral response	Strong/rapid bilateral response
Forelimb use	Severe bilateral deficits	Severe unilateral deficits	Slight deficits (unilateral or bilateral)	None
Hind limb use	Severe bilateral deficits	Severe unilateral deficits	Slight deficits (unilateral or bilateral)	None

Score range is from 0 (maximum deficits) to 24 (no deficits). SAH indicates subarachnoid hemorrhage.

24 hours (before neurobehavior could be completed), 3 mice died before day 2 neurobehavior, and 2 mice died before day 3 neurobehavior. Animals not surviving until euthanasia were excluded from the longitudinal analysis (16 sham and 15 SAH animals were included in analysis). All animals surviving >24 hours were included in the analysis of neurobehavioral performance on day 1 (16 sham and 20 SAH animals were included in analysis). For the 7-day study, one mouse in the SAH group died within 1 hour after SAH and another mouse died on day 2, so these data are not included in the longitudinal analysis (8 sham and 8 SAH animals were included in analysis). Data S1 contains all statistical reports (ie, exact *P* values, provided in Tables S1 through S8) as well as additional experimental results.

Modified Bederson Score

For the 3-day longitudinal study assessed using the modified Bederson Score, mice experiencing SAH perform significantly worse compared with sham mice on days 1 and 2 after injury (Figure 1A, Table 10). On day 3 after SAH, no significant difference between the sham and SAH animals was observed.

Katz Score

Using the Katz Score for the 3-day longitudinal study, SAH animals performed significantly worse that sham animals on all 3 days after SAH (Figure 1B, Table 10).

Garcia Neuroscore

The behavioral performance of mice on the Garcia Neuroscore for the 3-day longitudinal study was statistically different between the sham and SAH groups on all 3 days (Figure 1C, Table 10).

Parra Neuroscore

The Parra Neuroscore for the 3-day longitudinal study was able to statistically distinguish between sham and SAH mice on all 3 days after SAH (Figure 1D, Table 10).

Sensitivity Analysis of Behavioral Scoring Systems

For the mice subjected to neurobehavioral testing on day 1 after SAH, the data for each scoring system were examined to determine the diagnostic accuracy via ROC curve analysis (Table 11). Of note is the area under the curve; a greater area under the curve indicates that the test is better at placing a subject into the correct classification. The modified Bederson and Katz Scores had areas under the curves of 0.7313 and 0.8781, respectively. The areas under the curves for the Garcia and Parra Neuroscores were 0.8875 and 0.9266, respectively.

Analysis of the Sensitivity of the Individual Subtests

For all animals surviving for behavioral testing on day 1, the individual subtests were analyzed for statistical significance between the sham and SAH mice (Figure 2). For the balance and coordination tests, all subtests identified statistically significant deficits between the sham and SAH mice ($P<0.01$). For the posture and strength tests, all 3 subtests were able to detect significant deficits. The ptosis and facial weakness subtests had $P=0.0047$ and $P=0.0036$, respectively, whereas the dyspnea subtest had a $P=0.0417$. For the reflex tests, the side stroking, vibrissae touch, visual, olfactory, and righting subtests were able to distinguish between sham and SAH animals ($P<0.01$), whereas no significant deficits were observed in the tactile ($P=0.0529$), postural ($P=0.1930$), and sound ($P=0.1965$) subtests. All 3 subtests for the limb use tests identified significant differences between sham and SAH mice ($P<0.002$).

Using ROC curves, the ability for each subtest to correctly identify each animal as either sham or injured (ie, SAH) was computed (Table 12). The subtests with an area under the curve of <0.70 were forepaw outstretching, dyspnea, tactile reflex, postural reflex, and sound reflex. Those with an area under the curve between 0.70 and 0.75 were lateral turning and walking. The subtests with areas under the curve between 0.75 and 0.80 were beam walking, ptosis, facial weakness, olfactory reflex, righting reflex, and limb extension. Finally, the subtests with an area under the curve >0.80 were spontaneous activity, climbing, balance, side stroking, vibrissae touch, visual reflex, forelimb use, and hind limb use.

Development of a New Composite Neuroscore

After ROC curve analysis of the individual subtests, we found that the Parra Neuroscore uses one subtest that is insensitive toward SAH injury (the tactile reflex test). This led us toward developing a new composite neuroscore specifically designed to detect functional deficits observed after SAH in mice. Using the ROC curve analysis of the individual subtests, we set a threshold (for the area under the curve) equal to 0.80. All the subtests that had an area under the curve >0.80 were combined to make the new composite neuroscore (Table 9).

Using the new composite neuroscore developed within this study, we first analyzed the behavioral data from the 3-day study (Figure 3, Table 13). On days 1 to 3 after SAH, significant differences were observed between the sham and SAH mice. ROC curve analysis of the new composite

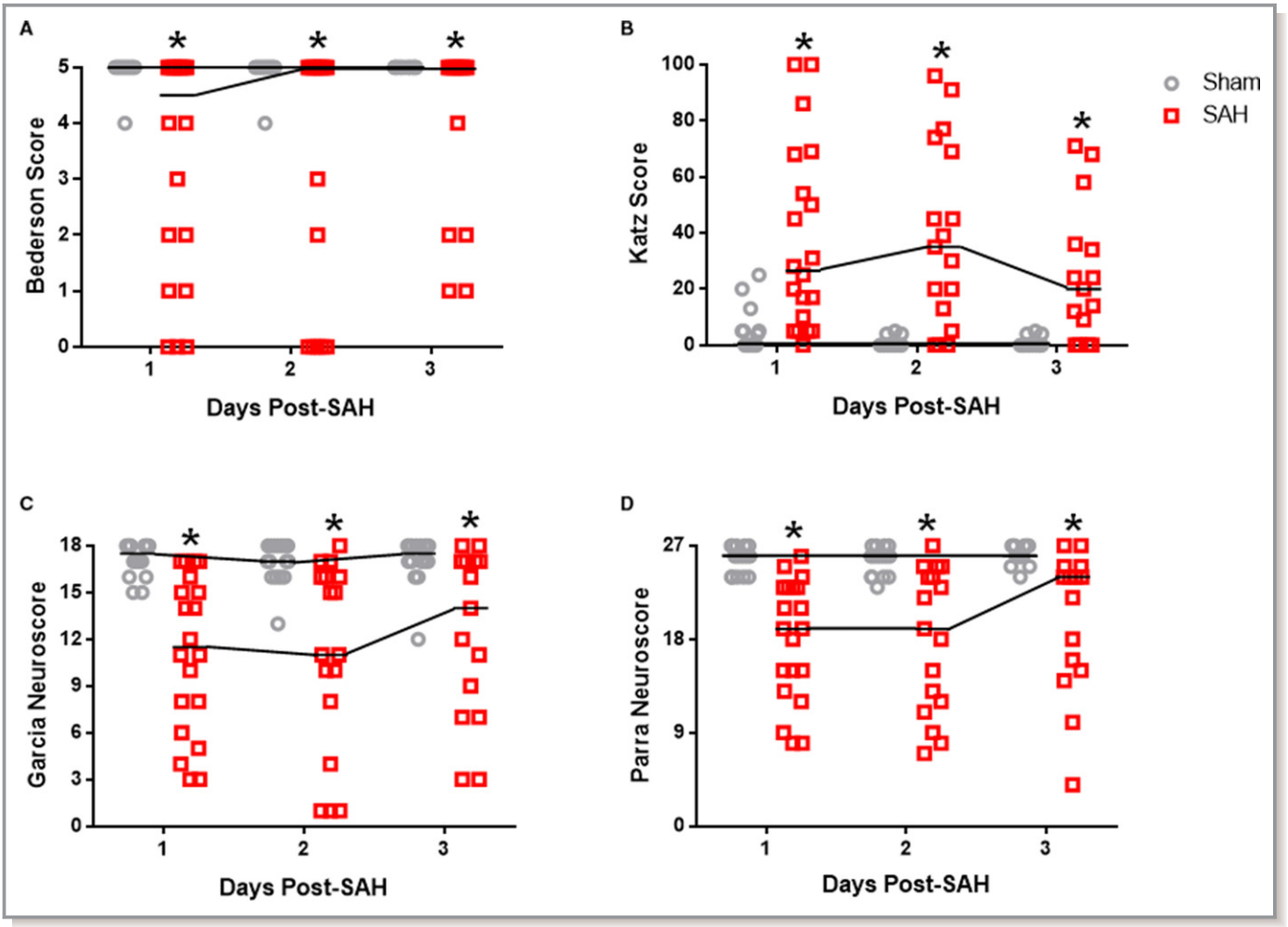

Figure 1. Neurobehavioral deficits on days 1 to 3 after subarachnoid hemorrhage (SAH). **A**, Modified Bederson Score. **B**, Katz Score. **C**, Garcia Neuroscore. **D**, Parra Neuroscore. Sham, n=16; SAH, n=15 to 20. Analyzed with Scheirer-Ray-Hare tests with Bonferroni post hoc tests. **P*<0.05 between sham and SAH at the indicated time point.

neuroscore measured the area under the curve to be 0.9953 (Table 11).

Finally, as an internal validation of our composite neuroscore, we performed a 7-day study on a separate cohort of mice (sham, n=8; SAH, n=10) (Table 14). Neither the modified Bederson nor Katz Score detected any significance difference in the neurobehavioral performance of sham versus SAH mice (Figure 4A and 4B). The Garcia Neuroscore detected significant differences between the 2 groups on days 2 and 3 after SAH, but not days 1, 5, and 7 (Figure 4C). The Parra Neuroscore found that SAH mice had significantly more deficits than sham animals on days 1 and 2 after SAH, but not any other day (Figure 4D). On days 1 to 3 and 5 after SAH, significant differences were observed between the sham and SAH mice for the developed composite neuroscore; deficits were not significantly different on day 7 (Figure 4E). The trends of individual mice are plotted on Figures S2 through S6.

ROC Curve Analysis of the Scoring Systems

To determine which neuroscore was the most useful for SAH mice studies, we analyzed the differences between the ROC curves for the Garcia Neuroscore, Parra Neuroscore, and our neuroscore (Table 15, Table S6). The Parra Neuroscore is not significantly more diagnostically accurate for SAH deficits than the Garcia Neuroscore (*P*=0.1062). However, the developed composite neuroscore is significantly more diagnostically accurate for SAH injury than both the Garcia Neuroscore (*P*=0.0121) and the Parra Neuroscore (*P*=0.0241).

Sample Size Estimations for Each of the Scoring Systems

Although a functional test may be diagnostically accurate for the injury studied, sample sizes are critical to every single study; and poor estimation of sample sizes required to test a

Table 10. Means and SDs for the 3-Day Study (Figure 1)

Scoring System	Mean	SD
Modified Bederson Score		
Day 1		
Sham	4.923	0.2774
SAH	3.350	1.568
Day 2		
Sham	4.923	0.2774
SAH	3.667	2.093
Day 3		
Sham	5.000	0.000
SAH	4.000	1.604
Katz Score		
Day 1		
Sham	5.231	8.604
SAH	28.27	23.51
Day 2		
Sham	1.000	1.915
SAH	24.67	24.39
Day 3		
Sham	1.000	1.915
SAH	16.92	1.115
Garcia Neuroscore		
Day 1		
Sham	16.92	1.124
SAH	11.15	5.008
Day 2		
Sham	16.69	1.377
SAH	11.73	5.849
Day 3		
Sham	16.85	1.625
SAH	12.40	5.422
Parra Neuroscore		
Day 1		
Sham	25.56	1.315
SAH	19.60	5.795
Day 2		
Sham	25.44	1.365
SAH	19.13	6.589
Day 3		
Sham	26.19	0.9106
SAH	19.93	6.829

SAH indicates subarachnoid hemorrhage.

Table 11. ROC Curve Analysis for the Neurobehavioral Scoring Systems

Scoring System	AUC (SE)	*P* Value
Modified Bederson Score	0.7313 (0.08392)	0.01852
Katz Score	0.8781 (0.05698)	0.0001
Garcia Neuroscore	0.8875 (0.05306)	<0.0001
Parra Neuroscore	0.9266 (0.04232)	<0.0001
Our composite neuroscore	0.9953 (0.00702)	<0.0001

Day 1 data from the 3-day study were used for analysis. AUC indicates area under the curve; ROC, receiver operating characteristic.

hypothesis in a particular study reduces the value of that study (ie, the study may be underpowered). Thus, we performed sample size estimations for each of the scoring systems used in this study on the basis of the functional performance of mice after SAH. Sample sizes were estimated for a variety of conditions: 2-group comparison (Mann-Whitney test) and 3-, 4-, and 5-group comparisons (Kruskal-Wallis test); most studies make statistical comparisons for these different conditions. The sample sizes required by the scoring systems are as follows: modified Bederson Score>Katz Score>Garcia Neuroscore>Parra Neuroscore>our composite neuroscore (Table 16). The sample sizes needed for the Parra Neuroscore are moderate, at n=8 to 12 per group, but the sample sizes needed to test for significance using our developed composite neuroscore are n=6 to 8 per group.

Finally, we estimated the sample size required to test for statistical significance in a longitudinal study (2 days) for 2 groups (sham and SAH) with the method of Liu and Liang,[18] using the mean differences, SDs of the populations, a desired power of 0.8, and an α of 0.05. Adjusting for the use of a nonparametric 2-way ANOVA test,[19] the sample sizes required are as follows: modified Bederson Score, n=32; Katz Score, n=23; Garcia Neuroscore, n=16; Parra Neuroscore, n=13; and our composite neuroscore, n=8.

Discussion

Herein, we developed a new composite neuroscore for measuring the functional performance of mice after SAH. Although various neurobehavioral testing schemes exist, none have been specifically designed nor validated for diagnostic accuracy to detect deficits after SAH. To our knowledge, this is the first SAH study to perform an analysis of the diagnostic accuracy for the modified Bederson Score, Katz Score, Garcia Neuroscore, and Parra Neuroscore. Furthermore, this is also the first study to have investigated the utility of the various subtests for SAH injury in mice. The shortcomings of the

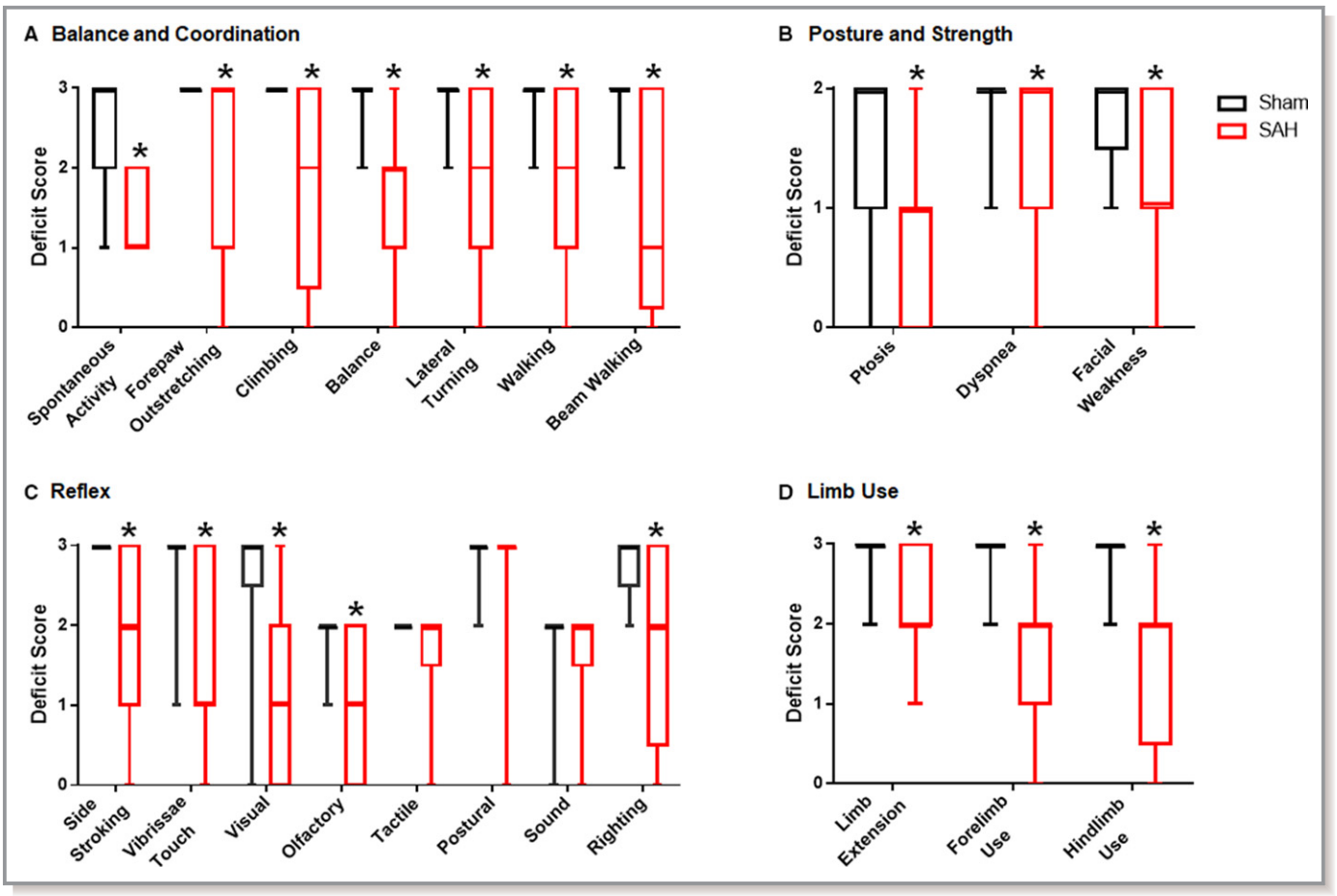

Figure 2. Functional deficits assessed by the individual subtests on day 1 after subarachnoid hemorrhage (SAH). **A**, Coordination and balance tests. **B,** Posture and strength tests. **C**, Reflex tests. **D**, Forelimb tests. Sham, n=16; SAH, n=20. Analyzed with Mann-Whitney tests. *P<0.05 vs sham for the indicated subtest.

preexisting scoring systems for SAH mice led to the development of a new composite score (Table 9). This composite score was specifically designed using individual subtests that are diagnostically accurate for SAH. Finally, this is also the first study to perform a sample size estimation comparison between the various scoring systems.

Within this study, we found that the modified Bederson Score is not an adequate scoring system for SAH in mice. Although both the Katz Score and the Garcia Neuroscore are somewhat diagnostically accurate for SAH injury (ie, have areas under the curves of 0.80–0.90), the Parra Neuroscore is a slightly better choice (because of the greater area under the curve). However, the developed composite neuroscore has an even greater ROC area under the curve and is even more diagnostically accurate for SAH injury than the Parra Neuroscore. The interoperator variation for the developed composite score is similar to the interoperator variations for the Garcia and Parra Neuroscores (all interoperator variations were >0.92 for the neurobehavioral data on day 1, Table 17).

Modified Bederson Score

Although the modified Bederson Score has never been used to assess injury after SAH in mice, it has been used in several rat models of SAH. Despite being used in rat SAH studies, the modified Bederson Score does not seem to be a suitable scoring system for identifying behavioral impairment after SAH in rats. The study by Bederson et al did not detect any functional deficits in rats on days 1 and 2 after SAH.[26] Similarly, the study by Bermueller et al observed no difference in the function of untreated and treated SAH rats for days 1 to 7 after SAH, despite positive findings in intracranial pressure reduction and neuronal survival for the treatment tested.[8] Finally, Thal et al did not observe any differences in the modified Bederson Score of untreated and treated rats on days 1 to 7 after SAH.[9] Furthermore, Thal et al[9] compared several other behavioral tests that were able to identify significant differences between untreated and treated animals after SAH, suggesting that the modified Bederson Score may not be applicable to SAH injury. The only positive study for use of the

Table 12. ROC Curve Analysis for the Individual Subtests

Subtest	AUC (SE)	*P* Value
Spontaneous activity	0.8277 (0.06900)	**0.0006**
Forepaw outstretching	0.6905 (0.08569)	**0.0460**
Climbing	0.8333 (0.06814)	**0.0005**
Balance	0.8782 (0.05749)	**<0.0001**
Lateral turning	0.7206 (0.08475)	**0.0223**
Walking	0.7451 (0.08021)	**0.0102**
Beam walking	0.7985 (0.07483)	**0.0020**
Ptosis	0.7647 (0.07947)	**0.0056**
Dyspnea	0.6401 (0.09032)	0.1422
Facial weakness	0.7563 (0.07961)	**0.0073**
Side stroking	0.8333 (0.06814)	**0.0005**
Vibrissae touch	0.8249 (0.07056)	**0.0007**
Visual reflex	0.8347 (0.07037)	**0.0005**
Olfactory reflex	0.7661 (0.07766)	**0.0053**
Tactile reflex	0.6190 (0.09082)	0.2122
Postural reflex	0.5714 (0.09321)	0.4541
Sound reflex	0.5882 (0.09273)	0.3551
Righting reflex	0.7899 (0.07373)	**0.0024**
Limb extension	0.7605 (0.07868)	**0.0063**
Forelimb use	0.8543 (0.06273)	**0.0002**
Hind limb use	0.8585 (0.06184)	**0.0002**

The day 1 data from the 3-day study were used for analysis. Boldfacing indicates $P<0.05$ between the sham and SAH groups on day 1 post-SAH. AUC indicates area under the curve; ROC, receiver operating characteristic.

modified Bederson Score for SAH in rats is that of Hockel et al, which reported significant deficits between untreated and treated SAH rats (n=10 per group) 1 and 2 days after injury, but

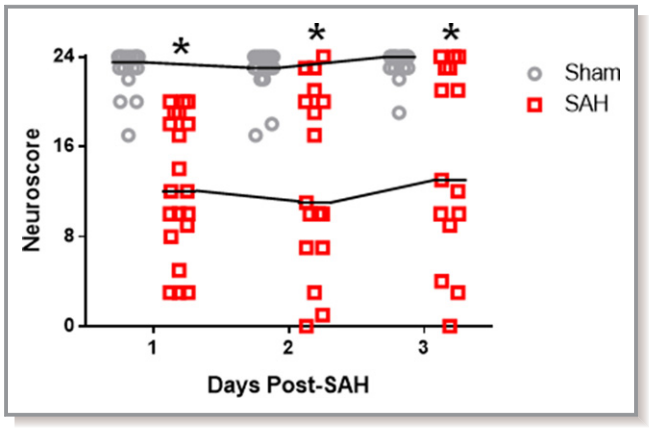

Figure 3. Assessing functional deficits on days 1 to 3 after subarachnoid hemorrhage (SAH) using the developed composite neuroscore. Sham, n=16; SAH, n=15 to 20. Analyzed with Scheirer-Ray-Hare test with Bonferroni post hoc test. *$P<0.05$ between sham and SAH at the indicated time point.

Table 13. Means and SDs for the 3-Day Study Using the Developed Composite Neuroscore (Figure 3)

Study Day (in 3-Day Study)	Mean	SD
Day 1		
Sham	22.81	1.223
SAH	12.70	6.105
Day 2		
Sham	22.63	2.125
SAH	13.29	8.099
Day 3		
Sham	23.31	1.302
SAH	14.73	8.581

SAH indicates subarachnoid hemorrhage.

no difference on days 3 to 7.[27] The modified Bederson Score was developed by modified Bederson et al for detecting injury after ischemic stroke, and thus the modified Bederson Score may be more suitable for revealing deficits from either unilateral injury or large ischemic damage rather than SAH injury.[28]

In our 3-day study, the modified Bederson Score was able to detect significant differences between sham and SAH mice on all 3 days, but no differences were found on days 1 to 3 in the 7-day study. Although this seems contradictory, it is

Table 14. Means and SDs for the 7-Day Study (Figure 4)

	Garcia Neuroscore		Parra Neuroscore		Our Composite Neuroscore	
Study Day	Mean	SD	Mean	SD	Mean	SD
Day 1						
Sham	17.50	0.7559	26.38	0.9161	23.88	0.3536
SAH	13.33	4.062	21.78	3.734	16.33	4.500
Day 2						
Sham	17.63	0.5175	26.38	0.9161	23.50	0.7559
SAH	13.63	3.462	22.25	4.234	17.13	6.578
Day 3						
Sham	17.75	0.4629	26.63	0.7440	23.63	0.7440
SAH	13.88	3.137	22.63	3.503	18.63	4.749
Day 5						
Sham	17.75	0.4629	26.63	0.7440	23.88	0.3536
SAH	14.88	2.997	24.00	2.726	19.63	4.274
Day 7						
Sham	17.88	0.3536	26.38	1.408	23.88	0.3536
SAH	15.50	3.024	24.25	3.327	20.75	3.991

SAH indicates subarachnoid hemorrhage.

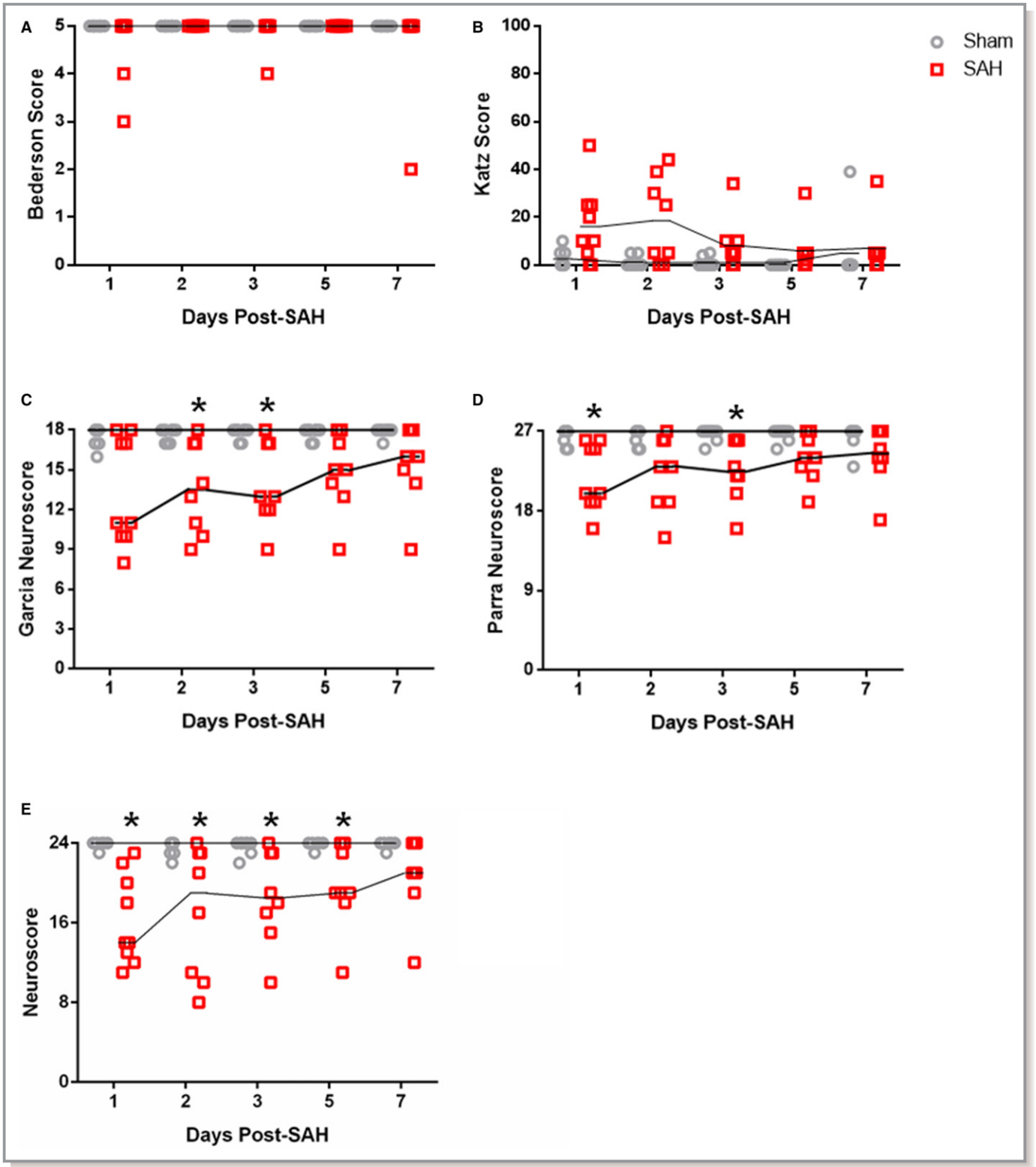

Figure 4. Neurobehavioral deficits on days 1 to 7 after subarachnoid hemorrhage (SAH) in mice. **A**, Modified Bederson Score. **B**, Katz Score. **C**, Garcia Neuroscore. **D**, Parra Neuroscore. **E**, Developed composite neuroscore. Sham, n=8; SAH, n=8 to 9. Analyzed with Scheirer-Ray-Hare test with Bonferroni post hoc test. **P*<0.05 between sham and SAH at the indicated time point.

attributable to the sample sizes for the 2 studies; the 3-day study was powered (ie, n=16–21 per group) to detect differences between sham and SAH mice for all the scoring systems, whereas the 7-day study was powered to detect differences for the Garcia Neuroscore, Parra Neuroscore, and the developed neuroscore (ie, n=8–10 per group). Our power

Table 15. Comparison of the ROC Curves for the Garcia Neuroscore, Parra Neuroscore, and the Developed Composite Neuroscore

Scoring System	Absolute Difference in AUC	SE_1; SE_2	*r*	*z* Score	*P* Value
Parra Neuroscore vs Garcia Neuroscore	0.0391	0.05306; 0.04232	0.807	1.248	0.1062
Our composite neuroscore vs Garcia Neuroscore	0.1078	0.05306; 0.00702	0.78	2.256	0.0121*
Our composite neuroscore vs Parra Neuroscore	0.0687	0.04232; 0.00702	0.84	1.877	0.0241*

SE_1 and SE_2 are the SEs of the AUCs (from ROC curve analysis) for the first and second scoring system being analyzed, respectively. The *r* value is the estimated correlation between the 2 scoring systems (obtained from the table provided by Hanley and McNeil[25]). AUC indicates area under the curve; ROC, receiver operating characteristic.
*Boldfacing indicates that there is a significant difference between the 2 scoring systems.

analysis of the modified Bederson Score data indicated that >20 mice in each group were needed to reach statistical significance, which is much more than the 10 mice allocated into the groups for the 7-day study (Table 16).

Katz Score

The Katz Score has been used in 5 rat SAH studies with mixed findings. In a study by Zausinger et al, the Katz Score was able to detect significant differences between treated and untreated rats only on the first day after SAH, although the study was conducted for 7 days.[29] In a follow-up study, Bermueller et al[8] used the Katz Score to test for significant functional differences in treated and untreated SAH rats (n=15/group) on days 1, 3, and 7 after SAH. Although the treatment reduced intracranial pressure and promoted neuron survival, no difference was found between the behavior of treated and untreated SAH rats using the Katz Score.[8] Although the first 2 studies only investigated functional deficits between treated and untreated rats, the study by Scholler et al tested for differences between sham and SAH rats at 6, 24, 48, and 72 hours after ictus.[10] This study observed significant differences between sham and SAH rats at 6 and 24 hours after SAH, but not on days 2 or 3. Finally, Thal et al observed functional deficits using the Katz Score on day 1 after SAH (SAH untreated versus SAH treated rats), but not on days 2 to 7.[9,30]

The studies by Zausinger et al[29] and Thal et al[9] (2008) tested for statistical significance between 3 groups, with 20 rats per group allocated, which is above our estimated sample size for 3 groups of n=15 per group. The study by Bermueller et al[8] allocated 15 rats per group and 4 groups, which is slightly lower than our estimated sample size (n=17/group). In the 2-group studies by Scholler et al[10] and Thal et al[30] (2009), 9 and 7 rats, respectively, were assigned to each group. Although these sample sizes are less than our estimated sample size for 2 groups, the findings may be attributed to no rat in the sham group having any deficits (ie, mean=0, SD=0). The findings of the current study suggest that if the correct number of animals is allocated into each group, the Katz Score is a viable option for examining functional deficits after SAH in rodents.

Similar to the modified Bederson Score, the Katz Score was adequate to identify significant differences between sham and SAH animals on all 3 days in the 3-day studies but failed to find differences for days 1 to 3 in the 7-day study. Again, this discrepancy is attributable to the way the 2 studies were powered (see Table 16 for power analysis).

Table 16. Sample Size Estimations for a Single Time Point Study

	No. of Groups			
Scoring System	2 (Mann-Whitney)	3 (Kruskal-Wallis)	4 (Kruskal-Wallis)	5 (Kruskal-Wallis)
Modified Bederson Score	17	21	24	25
Katz Score	13	15	17	19
Garcia Neuroscore	9	12	13	14
Parra Neuroscore	8	9	10	12
Our composite neuroscore	6	7	7	8

Sample sizes (ie, number of mice required in each group) for each neurobehavioral scoring system were estimated by SigmaPlot using the mean differences, SDs of the populations, a desired power of 0.8, and an α of 0.05. Because the sample sizes estimated by SigmaPlot reflect the assumptions of a 1-way ANOVA, each sample size was increased by 15% to estimate the sample sizes required to test for statistical significance using a nonparametric test (ie, 1-way ANOVA on ranks or Mann-Whitney).[19]

Table 17. Interoperator Variation

Scoring System	Tester 1 vs Tester 2	Tester 1 vs Tester 3	Tester 2 vs Tester 3
Modified Bederson Score	0.9364 (0.05116)	0.9134 (0.05257)	0.9357 (0.06001)
Katz Score	0.8818 (0.05856)	0.8175 (0.05329)	0.8823 (0.05856)
Garcia Neuroscore	0.9597 (0.01389)	0.9226 (0.03770)	0.9291 (0.02940)
Parra Neuroscore	0.9748 (0.00927)	0.9451 (0.02807)	0.9532 (0.02153)
Our composite neuroscore	0.9604 (0.01473)	0.9258 (0.03026)	0.9408 (0.02853)

Data are given as κ (SE). Data from day 1 of the 3-day study are used.

Garcia Neuroscore

By far, the Garcia Neuroscore has been much more used to assess functional deficits in rodents after SAH. The Garcia Neuroscore (and its modifications, which include addition of 1 more subtest) is reported to distinguish between sham and SAH mice in several studies.[31–33] Sozen et al observed differences between sham and SAH mice at 1 and 3 days after ictus using n=17 to 20 mice per group.[33] Similarly, Fujimoto et al reported functional deficits between uninjured and injured mice on days 1 and 2 after SAH (n=16–24 mice per group).[31] The first 2 studies used sample sizes much higher than required to test for statistical significance. Liu et al reported that mice after SAH (via endovascular perforation) performed significantly worse on the Garcia Neuroscore compared with sham mice.[32] This study allocated 10 mice into the sham group and 15 animals into the SAH group (of which 11 survived) to test for statistical significance in the Garcia Neuroscore. The authors ran their Kruskal-Wallis test on 4 experimental groups. Their sample size used (n=8–12 per group) is close to the sample size estimated for a 4-group ANOVA on ranks (ie, Kruskal-Wallis test) based on the data obtained in the current study (required sample size to test for significance was n=13 [Table 16]).

Much of the use of the Garcia Neuroscore has been performed in rats after SAH. Overall, SAH in rats after endovascular perforation leads to statistically significant differences in the functional performance of untreated SAH rats and sham rats on days 1 and 3 after injury.[11–13,15,34,35] Sugawara et al observed that moderate and severe SAH in rats leads to significant functional deficits 1 day after ictus (endovascular perforation model).[16] In addition, the authors reported on the interindividual variation of the Garcia Neuroscore for SAH rats and observed good correlation between the 2 scorers for identifying the presence or lack of deficits.

In general, the original 6 sensorimotor subtests proposed by Garcia et al[23] have been used for assessing functional deficits after SAH in rats. Despite the lack of sensitivity for the forepaw outstretching subtest, most investigators have observed statistically significant deficits after SAH compared with sham rats.[11–13,15,34,35] Other researchers have used a modified Garcia Neuroscore combining 7 subtests,[36,37] which sometimes includes the beam walking subtest.[38–40]

The Garcia Neuroscore (and its modified version) has been successful for identifying functional deficits after SAH in rats, and the current study (as well as others[31–33]) indicates that this neuroscoring system may also be useful for assessing deficits after SAH in mice. However, the findings from our study and those of Liu et al indicate that a larger sample size is required to test for statistical significance in the mouse SAH model[31–33] compared with the rat model.[36–38,40] One reason for this difference between endovascular perforation in rats versus mice is that rats are prone to larger hemorrhages, as well as slower hemorrhage clearance. Specifically, large subarachnoid hemorrhages can be observed and graded in rats after SAH,[16] but in mice this is more difficult because of smaller hemorrhages.[17,41–43] In addition, the hemorrhage in rats can be observed and graded 3 days after SAH[14,44] and may last as long as 7 days after injury,[37] whereas in the mouse, the hemorrhage is typically cleared by 2 to 4 days after SAH.[42,43]

Furthermore, inclusion of the beam balance score in the Garcia Neuroscore may provide added diagnostic accuracy to the modified Garcia Neuroscore. In the current study, we did not include the beam balance test into the Garcia Neuroscore. However, the sensitivity of the beam balance test for assessing functional deficits after SAH in mice (area under the curve of 0.799) suggests that modifying the Garcia Neuroscore to include the beam walking test may increase its diagnostic accuracy. The study by Liu et al used the endovascular perforation mouse model and observed significant functional deficits in the beam balance score, as well as the Garcia Neuroscore, further suggesting that there might be added sensitivity if the beam balance score is incorporated into the Garcia Neuroscore.[32] However, because the Garcia Neuroscore is significantly less diagnostically accurate for SAH injury than the developed neuroscore (because it includes the forepaw outstretching subtest, which is insensitive), it is unlikely that including the beam walking test in the Garcia Neuroscore will make it better than the composite neuroscore developed in this study.

Parra Neuroscore

To date, the Parra Neuroscore has been used in 5 studies.[17,45–48] Although it is a composite score, it combines several tests that are distinct from those included in the Garcia Neuroscore. The forepaw stretching test (part of the Garcia Neuroscore), which we found was not sensitive for SAH injured, is not included in the Parra Neuroscore. Rather, the Parra Neuroscore makes use of the visual reflex test, which is highly diagnostically accurate to SAH injury, and the beam walking and olfactory tests, which are moderately diagnostically accurate to SAH deficits. However, also included, which likely diminishes the diagnostic accuracy of the Parra Neuroscore, is the insensitive tactile reflex subtest. In the original study, the Parra Neuroscore was capable of detecting behavioral differences between sham and SAH mice 3 days after SAH.[17] The animal number distributed to each group was n=7 to 8, which agrees with our sample size estimation of n=8 per group for 2-group analysis. In 3 more recent articles, the studies by Vellimana et al, Han et al, and Wu et al found that the Parra Neuroscore was able to distinguish between sham and injured animals (as well as to evaluate treatment effects) using n=11 to 16,[47] n=13 to 20,[17] and n=15 to 28[48] mice, respectively. These 3 studies had sample sizes well above that needed to test for statistical significance. One negative study using the Parra Neuroscore was that of Tait et al.[46] The authors were unable to detect any significant differences between sham and SAH mice at either 6 or 24 hours after injury.[46] However, this is likely attributable to being slightly underpowered; only 3 mice were included in the sham group and 10 to 12 mice were included in the injured groups.

Other Neuroscoring Systems

There also exist several other composite neuroscores that combine subtesting of sensory-motor function. Specifically used in mice after SAH are the neuroscores used by the Laskowitz' group,[49–52] Plesnila's group,[41,53] McGirt et al,[54,55] and Neulen et al.[56] Two other scoring systems used for murine SAH are the 3-point scale used by Tamargo's group[57,58] and the SHIRPA (SmithKline Beecham Pharmaceuticals; Harwell, MRC Mouse Genome Centre and Mammalian Genetics Unit; Imperial College School of Medicine at St Mary's; Royal London Hospital, St Bartholomew's and the Royal London School of Medicine; Phenotype Assessment) score.[59,60]

The Laskowitz Neuroscore combines the spontaneous activity, limb extension, climbing, balance and coordination, side stroking, vibrissae touch, visual, and tactile subtests.[49–52] Of these subtests in the neuroscore by Laskowitz' group, the findings of the current study indicate that all expect the tactile test is sensitive and specific for SAH. Thus, the diagnostic accuracy of their neuroscore may be reduced compared with the composite neuroscore developed within because of the inclusion of the tactile subtest.

Another neuroscoring system developed is that of Plesnila's group, which has been used for the mouse model of SAH via endovascular perforation.[41,53] The developed neuroscore combines tests of reflexes (grasping, righting, and falling), coordination (head orientation and circling), and general behaviors (spontaneous activity, fur appearance, nibbling, and flight) into a score that ranges from 0 (best performance) to 33 (worst performance). This group has demonstrated that their neuroscore is sensitive enough to detect significant deficits in SAH mice compared with sham mice for up to 7 days with a sample size of 6 to 9 mice per group (analyzed using Kruskal-Wallis ANOVA on ranks[41] or the Friedman test[53]). The positive findings by Plesnila's group using this scoring method,[41,53] as well as unambiguous scoring criterion, warrant this neuroscore be investigated in future studies.

McGirt et al[54,55] used 2 different composite neuroscoring systems to test for functional deficits after SAH in mice. Specifically, the authors used a 9- to 39-point scoring system[55] and a 5- to 27-point scoring method.[54] Both scoring schemes were acceptable in identifying behavioral deficits after SAH in mice. Although the specific scoring criteria for each subtest are not clear, the authors used sample sizes of n=14 to 17 per group to test for statistical significance.[54,55] The 5- to 27-point scoring method uses the same subtests as the Parra Neuroscore. However, because the exact scoring criterion for the 2 methods is missing, it is difficult for either of these scoring methods to be used by others.

The 3-point scale used by Tamargo's group has been used in mice subjected to SAH via autologous blood injected into the cisternal magna.[57,58] This scoring method observes 3 behaviors (namely, posture, grooming, and ambulation). Each behavior is scored either 0 (deficits) or 1 (no deficits), with a maximum score of 3 (uninjured) and a minimum score of 0 (severely injured). This test was performed by 2 independent observers, and the scores were averaged. This test has yet to be applied to the other SAH mouse models, as well as be evaluated for testing for deficits at time points >1 day.

One more interesting scoring system is the SHIRPA score.[59,60] To date, the SHIRPA score has been used in a mouse SAH study (specifically, the autologous blood injection model in the prechiasmatic space). Boettinger et al observed that mice with moderate to severe SAH (induced using 100–120 μL of blood) presented with significant functional deficits for body position (4 subtests) and motor behavior (10 subtests) and minor deficits in spontaneous activity (single test) and gait (single test) on days 1 and 2 after SAH.[59] Herein, this scoring system was not specifically used, and because the exact details about each individual score for the subtests are not clear, it is difficult to assess the utility of the SHIRPA score for determining behavioral deficits after SAH in mice.

Finally, although not developed for SAH, the neuroscore by Feldman et al measures mobility, reflexes, behavior, and function[61] and has been highlighted as a potentially useful scoring system for SAH in rodents because of its inclusion of motor and sensory tests, as well as beam walking and mobility.[6] A modification of this scoring system has been successfully applied to rats subjected to SAH; Yatsushige et al observed modest differences in the functional deficits of untreated versus treated rats after subarachnoid hemorrhage, but no comparison was made between the functional behavior of SAH rats and sham rats.[62] Additional studies need to be undertaken to determine the utility of the Feldman Neuroscore for detecting functional deficits after SAH.

Toward a Standardized Neuroscore for SAH Mice Studies

To our knowledge, this is the first study to investigate the utility and diagnostic accuracy of various neurobehavioral scoring systems for assessing deficits after SAH in rodents. The current study used mice, but the findings likely apply to rat studies (although specific sample size estimations may be different). Of the previously published scoring systems, the combined findings of literature and this study cannot recommend that the modified Bederson Score or the Katz Score be used for rodent SAH studies. This is because large sample sizes are required to test for statistical significance. As for the Garcia Neuroscore, although the sample size needed is modest, the ROC curve analysis suggests that the Parra Neuroscore is a better choice. But, because the Parra Neuroscore incorporates the insensitive tactile reflex subtest, we argue that it is still not the optimal neuroscoring system, and therefore we developed a new composite neuroscore using only highly diagnostically accurate subtests. The developed composite neuroscore is significantly more diagnostically accurate than both the Garcia Neuroscore (P=0.0121) and the Parra Neuroscore (P=0.0241). The findings of the current study suggest that the most appropriate scoring system for testing for functional deficits after SAH in mice is the new composite neuroscore developed within. The new neuroscore was demonstrated to have interoperator variation similar to the Garcia and Parra Neuroscores, and because several of the subtests overlap with the Garcia and Parra Neuroscores, the new neuroscore can be quickly adopted by other preclinical SAH groups without any new costs or training.

Limitations and Future Studies

The current study has several limitations that need to be addressed in future studies to fully understand the application of composite neuroscoring systems to experimental SAH studies. First, this study did not investigate all the various composite scoring tests. In future studies, we can investigate other scoring methods and compare them with the developed neuroscore. Second, the recommendations from the round-table committees argue for testing various species, different strains, both sexes, as well as age, and even comorbidities. We chose young C57 male mice specifically because they are the most widely used. However, these results may not apply to other strains of mice or even female C57BL/6 mice. We are continuing this study in C57BL/6 mice using aged males (10 and 18 months old), females (4, 10, and 18 months old), and also a blood injection model of SAH. Our plan is to publish our findings in future articles for these groups. With respect to the other SAH models, the other 2 primary models of SAH in mice are the autologous blood injection models into the prechiasmatic space and the cisterna magna. These models should be investigated for diagnostic accuracy for the various scoring methods in future studies. In addition, the diagnostic accuracy for these composite neuroscores may be different between mice and rats. Thus, although the developed neuroscore is the most appropriate for testing mice subjected to SAH via endovascular perforation, this may not hold true for other SAH models or for endovascular perforation in rats. Third, numerous other functional tests exist for evaluating functional behavior in mice: rotarod, Morris water maze, T-maze, corner turn test, and forelimb placement test, just to name a few. The reviews of Jeon et al[6] and Turan et al[7] are excellent reads for the current findings of these and other neurobehavior tests. Many of these other tests require unique equipment/setups to test for specific functions, whereas composite neuroscores test several general behaviors and are typically more economical and feasible to perform.

Conclusion

Herein, a composite neuroscore was developed for evaluating behavioral deficits in mice subjected to SAH. The new composite neuroscore was compared with several other functional scoring systems, and our neuroscore had greater diagnostic accuracy for SAH injury in mice. The results of this study suggest that the new composite neuroscore is more appropriate to use in mice studies of SAH than the modified Bederson or Katz Scores, or the Garcia or Parra Neuroscores.

Sources of Funding

Funding support was provided by a seed grant provided by The Vivian L. Smith Department of Neurosurgery at the University of Texas Health Science Center at Houston (McBride).

Disclosures

None.

References

1. McDonald RJ, McDonald JS, Bida JP, Kallmes DF, Cloft HJ. Subarachnoid hemorrhage incidence in the United States does not vary with season or temperature. *Am J Neuroradiol*. 2012;33:1663–1668.
2. Stroke Therapy Academic Industry Roundtable (STAIR). Recommendations for standards regarding preclinical neuroprotective and restorative drug development. *Stroke*. 1999;30:2752–2758.
3. Suzuki H, Nakano F. To improve translational research in subarachnoid hemorrhage. *Transl Stroke Res*. 2018;9:1–3.
4. Lapchak PA, Zhang JH, Noble-Haeusslein LJ. Rigor guidelines: escalating stair and steps for effective translational research. *Transl Stroke Res*. 2013;4:279–285.
5. Stem Cell Therapies as an Emerging Paradigm in Stroke Participants. Stem Cell Therapies as an Emerging Paradigm in Stroke (STEPS): bridging basic and clinical science for cellular and neurogenic factor therapy in treating stroke. *Stroke*. 2009;40:510–515.
6. Jeon H, Ai J, Sabri M, Tariq A, Shang X, Chen G, Macdonald RL. Neurological and neurobehavioral assessment of experimental subarachnoid hemorrhage. *BMC Neurosci*. 2009;10:103.
7. Turan N, Miller BA, Heider RA, Nadeem M, Sayeed I, Stein DG, Pradilla G. Neurobehavioral testing in subarachnoid hemorrhage: a review of methods and current findings in rodents. *J Cereb Blood Flow Metab*. 2017;37:3461–3474.
8. Bermueller C, Thal SC, Plesnila N, Schmid-Elsaesser R, Kreimeier U, Zausinger S. Hypertonic fluid resuscitation from subarachnoid hemorrhage in rats: a comparison between small volume resuscitation and mannitol. *J Neurol Sci*. 2006;241:73–82.
9. Thal SC, Mebmer K, Schmid-Elsaesser R, Zausinger S. Neurological impairment in rats after subarachnoid hemorrhage—a comparison of functional tests. *J Neurol Sci*. 2008;268:150–159.
10. Scholler K, Trinkl A, Klopotowski M, Thal SC, Plesnila N, Trabold R, Hamann GF, Schmid-Elsaesser R, Zausinger S. Characterization of microvascular basal lamina damage and blood-brain barrier dysfunction following subarachnoid hemorrhage in rats. *Brain Res*. 2007;1142:237–246.
11. Park S, Yamaguchi M, Zhou C, Calvert JW, Tang J, Zhang JH. Neurovascular protection reduces early brain injury after subarachnoid hemorrhage. *Stroke*. 2004;35:2412–2417.
12. Cahill J, Calvert JW, Marcantonio S, Zhang JH. P53 may play an orchestrating role in apoptotic cell death after experimental subarachnoid hemorrhage. *Neurosurgery*. 2007;60:531–545; discussion 545.
13. Cahill J, Calvert JW, Solaroglu I, Zhang JH. Vasospasm and p53-induced apoptosis in an experimental model of subarchnoid hemorrhage. *Stroke*. 2006;37:1868–1874.
14. Enkhjargal B, McBride DW, Manaenko A, Reis C, Sakai Y, Tang J, Zhang JH. Intranasal administration of vitamin D attenuates blood-brain barrier disruption through endogenous upregulation of osteopontin and activation of CD44/P-gp glycosylation signaling after subarachnoid hemorrhage in rats. *J Cereb Blood Flow Metab*. 2017;37:2555–2566.
15. Ostrowski RP, Colohan AR, Zhang JH. Mechanisms of hyperbaric oxygen-induced neuroprotection in a rat model of subarachnoid hemorrhage. *J Cereb Blood Flow Metab*. 2005;25:554–571.
16. Sugawara T, Ayer R, Jadhav V, Zhang JH. A new grading system evaluating bleeding scale in filament perforation subarachnoid hemorrhage rat model. *J Neurosci Methods*. 2008;167:327–334.
17. Han BH, Vellimana AK, Zhou ML, Milner E, Zipfel GJ. Phosphodiesterase 5 inhibition attenuates cerebral vasospasm and improves functional recovery after experimental subarachnoid hemorrhage. *Neurosurgery*. 2012;70:178–86; discussion 186–7.
18. Liu G, Liang KY. Sample size calculations for studies with correlated observations. *Biometrics*. 1997;53:937–947.
19. Lehmann EL. *Nonparametrics: Statistical Methods Based on Ranks*. Springer, New York; 2006.
20. Aum DJ, Vellimana AK, Singh I, Milner E, Nelson JW, Han BH, Zipfel GJ. A novel fluorescent imaging technique for assessment of cerebral vasospasm after experimental subarachnoid hemorrhage. *Sci Rep*. 2017;7:9126.
21. McBride DW, Nowrangi D, Kaur H, Wu G, Huang L, Lekic T, Tang J, Zhang JH. A composite neurobehavioral test to evaluate acute functional deficits after cerebellar haemorrhage in rats. *J Cereb Blood Flow Metab*. 2018;38:433–446.
22. Katz L, Ebmeyer U, Safar P, Radovsky A, Neumar R. Outcome model of asphyxial cardiac arrest in rats. *J Cereb Blood Flow Metab*. 1995;15:1032–1039.
23. Garcia JH, Wagner S, Liu KF, Hu XJ. Neurological deficit and extent of neuronal necrosis attributable to middle cerebral artery occlusion in rats: statistical validation. *Stroke*. 1995;26:627–634; discussion 635.
24. Lang TA, Secic M. *How to Report Statistics in Medicine*. Philadelphia, PA. American College of Physicians; 2006.
25. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983;148:839–843.
26. Bederson JB, Germano IM, Guarino L. Cortical blood flow and cerebral perfusion pressure in a new noncraniotomy model of subarachnoid hemorrhage in the rat. *Stroke*. 1995;26:1086–1091; discussion 1091–1092.
27. Hockel K, Scholler K, Trabold T, Nussberger J, Plesnila N. Vasopressin V1a receptors mediate posthemorrhagic systemic hypertension thereby determining rebleeding rate and outcome after experimental subarachnoid hemorrhage. *Stroke*. 2012;43:227–232.
28. Bederson JB, Pitts LH, Tsuji M, Nishimura MC, Davis RL, Bartkowski H. Rat middle cerebral artery occlusion: evaluation of the model and development of a neurologic examination. *Stroke*. 1986;17:472–476.
29. Zausinger S, Thal SC, Kreimeier U, Messmer K, Schmid-Elsaesser R. Hypertonic fluid resuscitation from subarachnoid hemorrhage in rats. *Neurosurgery*. 2004;55:679–686; discussion 686–687.
30. Thal SC, Sporer S, Klopotowski M, Thal SE, Woitzik J, Schmid-Elsaesser R, Plesnila N, Zausinger S. Brain edema formation and neurological impairment after subarachnoid hemorrhage in rats: laboratory investigation. *J Neurosurg*. 2009;111:988–994.
31. Fujimoto M, Shiba M, Kawakita F, Liu L, Shimojo N, Imanaka-Yoshida K, Yoshida T, Suzuki H. Deficiency of tenascin-C and attenuation of blood-brain barrier disruption following experimental subarachnoid hemorrhage in mice. *J Neurosurg*. 2016;124:1693–1702.
32. Liu L, Fujimoto M, Nakano F, Nishikawa H, Okada T, Kawakita F, Imanaka-Yoshida K, Yoshida T, Suzuki H. Deficiency of tenascin-C alleviates neuronal apoptosis and neuroinflammation after experimental subarachnoid hemorrhage in mice. *Mol Neurobiol*. 2018;55:8346–8354.
33. Sozen T, Tsuchiyama R, Hasegawa Y, Suzuki H, Jadhav V, Nishizawa S, Zhang JH. Role of interleukin-1beta in early brain injury after subarachnoid hemorrhage in mice. *Stroke*. 2009;40:2519–2525.
34. Sugawara T, Ayer R, Jadhav V, Chen W, Tsubokawa T, Zhang JH. Simvastatin attenuation of cerebral vasospasm after subarachnoid hemorrhage in rats via increased phosphorylation of Akt and endothelial nitric oxide synthase. *J Neurosci Res*. 2008;86:3635–3643.
35. Gao C, Liu X, Liu W, Shi H, Zhao Z, Chen H, Zhao S. Anti-apoptotic and neuroprotective effects of tetramethylpyrazine following subarachnoid hemorrhage in rats. *Auton Neurosci*. 2008;141:22–30.
36. Guo Z, Hu Q, Xu L, Guo ZN, Ou Y, He Y, Yin C, Sun X, Tang J, Zhang JH. Lipoxin A4 reduces inflammation through formyl peptide receptor 2/p38 MAPK signaling pathway in subarachnoid hemorrhage rats. *Stroke*. 2016;47:490–497.
37. Liu F, Hu Q, Li B, Manaenko A, Chen Y, Tang J, Zhang JH. Recombinant milk fat globule-EGF factor-8 reduces oxidative stress via integrin beta3/nuclear factor erythroid 2-related factor 2/heme oxygenase pathway in subarachnoid hemorrhage rats. *Stroke*. 2014;45:3691–3697.
38. Chen S, Ma Q, Krafft PR, Hu Q, Rolland W II, Sherchan P, Zhang J, Tang J, Zhang JH. P2X7R/cryopyrin inflammasome axis inhibition reduces neuroinflammation after SAH. *Neurobiol Dis*. 2013;58:296–307.
39. Fujii M, Sherchan P, Krafft PR, Rolland WB, Soejima Y, Zhang JH. Cannabinoid type 2 receptor stimulation attenuates brain edema by reducing cerebral leukocyte infiltration following subarachnoid hemorrhage in rats. *J Neurol Sci*. 2014;342:101–106.
40. Tang J, Hu Q, Chen Y, Liu F, Zheng Y, Tang J, Zhang J, Zhang JH. Neuroprotective role of an N-acetyl serotonin derivative via activation of tropomyosin-related kinase receptor B after subarachnoid hemorrhage in a rat model. *Neurobiol Dis*. 2015;78:126–133.
41. Feiler S, Friedrich B, Scholler K, Thal SC, Plesnila N. Standardized induction of subarachnoid hemorrhage in mice by intracranial pressure monitoring. *J Neurosci Methods*. 2010;190:164–170.
42. Lin CL, Calisaneller T, Ukita N, Dumont AS, Kassell NF, Lee KS. A murine model of subarachnoid hemorrhage-induced cerebral vasospasm. *J Neurosci Methods*. 2003;123:89–97.
43. Muroi C, Fujioka M, Okuchi K, Fandino J, Keller E, Sakamoto Y, Mishima K, Iwasaki K, Fujiwara M. Filament perforation model for mouse subarachnoid hemorrhage: surgical-technical considerations. *Br J Neurosurg*. 2014;28:722–732.
44. Wu J, Zhang Y, Yang P, Enkhjargal B, Manaenko A, Tang J, Pearce WJ, Hartman R, Obenaus A, Chen G, Zhang JH. Recombinant osteopontin stabilizes smooth muscle cell phenotype via integrin receptor/integrin-linked kinase/Rac-1 pathway after subarachnoid hemorrhage in rats. *Stroke*. 2016;47:1319–1327.
45. Han BH, Vellimana AK, Zhou ML, Milner E, Zipfel GJ. Phosphodiesterase 5 inhibition attenuates cerebral vasospasm and improves functional recovery after experimental subarachnoid hemorrhage. *Neurosurgery*. 2012;70:178–186; discussion 186–187.

46. Tait MJ, Saadoun S, Bell BA, Verkman AS, Papadopoulos MC. Increased brain edema in aqp4-null mice in an experimental model of subarachnoid hemorrhage. *Neuroscience*. 2010;167:60–67.
47. Vellimana AK, Milner E, Azad TD, Harries MD, Zhou ML, Gidday JM, Han BH, Zipfel GJ. Endothelial nitric oxide synthase mediates endogenous protection against subarachnoid hemorrhage-induced cerebral vasospasm. *Stroke*. 2011;42:776–782.
48. Wu Y, Pang J, Peng J, Cao F, Vitek MP, Li F, Jiang Y, Sun X. An apoE-derived mimic peptide, COG1410, alleviates early brain injury via reducing apoptosis and neuroinflammation in a mouse model of subarachnoid hemorrhage. *Neurosci Lett*. 2016;627:92–99.
49. Gao J, Wang H, Sheng H, Lynch JR, Warner DS, Durham L, Vitek MP, Laskowitz DT. A novel apoE-derived therapeutic reduces vasospasm and improves outcome in a murine model of subarachnoid hemorrhage. *Neurocrit Care*. 2006;4:25–31.
50. Mesis RG, Wang H, Lombard FW, Yates R, Vitek MP, Borel CO, Warner DS, Laskowitz DT. Dissociation between vasospasm and functional improvement in a murine model of subarachnoid hemorrhage. *Neurosurg Focus*. 2006;21:E4.
51. Wang H, Gao J, Lassiter TF, McDonagh DL, Sheng H, Warner DS, Lynch JR, Laskowitz DT. Levetiracetam is neuroprotective in murine models of closed head injury and subarachnoid hemorrhage. *Neurocrit Care*. 2006;5:71–78.
52. Wang H, James ML, Venkatraman TN, Wilson LJ, Lyuboslavsky P, Myers SJ, Lascola CD, Laskowitz DR. pH-sensitive NMDA inhibitors improve outcome in a murine model of SAH. *Neurocrit Care*. 2014;20:119–131.
53. Terpolilli NA, Feiler S, Dienel A, Muller F, Heumos N, Friedrich B, Stover J, Thal S, Scholler K, Plesnila N. Nitric oxide inhalation reduces brain damage, prevents mortality, and improves neurological outcome after subarachnoid hemorrhage by resolving early pial microvasospasms. *J Cereb Blood Flow Metab*. 2016;36:2096–2107.
54. McGirt MJ, Lynch JR, Parra A, Sheng H, Pearlstein RD, Laskowitz DT, Pelligrino DA, Warner DS. Simvastatin increases endothelial nitric oxide synthase and ameliorates cerebral vasospasm resulting from subarachnoid hemorrhage. *Stroke*. 2002;33:2950–2956.
55. McGirt MJ, Parra A, Sheng H, Higuchi Y, Oury TD, Laskowitz DT, Pearlstein RD, Warner DS. Attenuation of cerebral vasospasm after subarachnoid hemorrhage in mice overexpressing extracellular superoxide dismutase. *Stroke*. 2002;33:2317–2323.
56. Neulen A, Meyer S, Kramer A, Pantel T, Kosterhon M, Kunzelmann S, Goetz H, Thal SC. Large vessel vasospasm is not associated with cerebral cortical hypoperfusion in a murine model of subarachnoid hemorrhage. *Transl Stroke Res*. 2018. Available at: https://doi.org/10.1007/s12975-018-0647-6. Accessed March 6, 2019.
57. Chaichana KL, Levy AP, Miller-Lotan R, Shakur S, Tamargo RJ. Haptoglobin 2-2 genotype determines chronic vasospasm after experimental subarachnoid hemorrhage. *Stroke*. 2007;38:3266–3271.
58. Momin EN, Schwab KE, Chaichana KL, Miller-Lotan R, Levy AP, Tamargo RJ. Controlled delivery of nitric oxide inhibits leukocyte migration and prevents vasospasm in haptoglobin 2-2 mice after subarachnoid hemorrhage. *Neurosurgery*. 2009;65:937–945; discussion 945.
59. Boettinger S, Kolk F, Broessner G, Helbok R, Pfausler B, Schmutzhard E, Beer R, Lackner P. Behavioral characterization of the anterior injection model of subarachnoid hemorrhage. *Behav Brain Res*. 2017;323:154–161.
60. Rogers DC, Fisher EM, Brown SD, Peters J, Hunter AJ, Martin JE. Behavioral and functional analysis of mouse phenotype: SHIRPA, a proposed protocol for comprehensive phenotype assessment. *Mamm Genome*. 1997;8:711–713.
61. Feldman Z, Gurevitch B, Artru AA, Oppenheim A, Shohami E, Reichenthal E, Shapira Y. Effect of magnesium given 1 hour after head trauma on brain edema and neurological outcome. *J Neurosurg*. 1996;85:131–137.
62. Yatsushige H, Ostrowski RP, Tsubokawa T, Colohan A, Zhang JH. Role of c-Jun N-terminal kinase in early brain injury after subarachnoid hemorrhage. *J Neurosci Res*. 2007;85:1436–1448.

Supplemental Material

Data S1.

0.70 used Spontaneous Activity, Climbing, Balance, Lateral Turning, Walking, Beam Walking, Ptosis, Facial Weakness, Side Stroking, Vibrissae Touch, Visual Reflex, Olfactory Reflex, Righting Reflex, Limb Extension, Forelimb Use, and Hindlimb Use.

The composite neuroscore using sub-tests which had areas under the curves of greater than 0.75 examined Spontaneous Activity, Climbing, Balance, Beam Walking, Ptosis, Facial Weakness, Side Stroking, Vibrissae Touch, Visual Reflex, Olfactory Reflex, Righting Reflex, Limb Extension, Forelimb Use, and Hindlimb Use.

Finally, the composite neuroscore based on the sub-tests with an area under the curve greater than 0.80 uses Spontaneous Activity, Climbing, Balance, Side Stroking, Vibrissae Touch, Visual Reflex, Forelimb Use, and Hindlimb Use.

After carefully evaluating the ROC curves for each of these combinations (using behavioral data from day 1 of the 3-day study, the most accurate and precise composite neuroscore developed was based on an individual sub-test threshold of 0.80. Briefly, the areas under the curves for the composite neuroscores based on sub-test thresholds of 0.70, 0.75, and 0.80 were 0.9313, 0.9313, and 0.9953, respectively. Therefore, the optimal composite neuroscoring system, based on the sub-tests investigated within this study was the composite neuroscore using a sub-test area under the curve threshold of 0.80.

We also developed a composite neuroscore based on a variable selection procedure through Lasso regression. This method identified the following sub-tests as the most important (frequency greater than 150, Figure S1): side stroking, beam walking, spontaneous activity, balance, forelimb use, olfactory, and vibrissae touch. Combining these sub-tests into a neuroscore yields an area under the curve for the ROC curve equal to 0.9688 (0.02388 error).

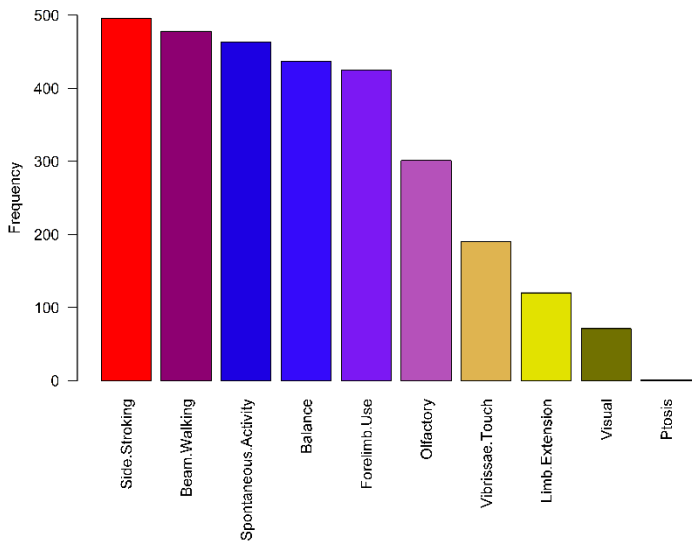

Figure S1. Variable Selection through Lasso Regression to Identify the Most Important Sub-tests for the Composite Neuroscore. All sub-tests were subjected to this process, but only the top 10 are shown for the frequency.

Therefore, based on the ROC curves for the various neuroscores, the optimal composite neuroscore for detecting SAH injury was the neuroscore based on a sub-test ROC curve threshold of greater than 0.80 (Table 9).

Neurobehavioral Performance over the First 7 Days Post-SAH

The performances of individual mice on each of the scoring systems for the 7-day study are shown in Figures S2-S6.

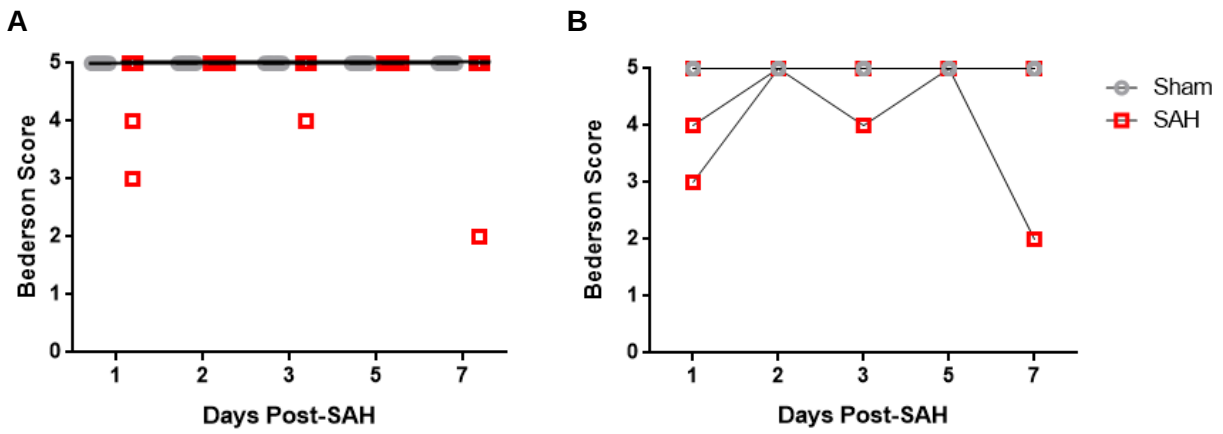

Figure S2. Performance on the Bederson Score for Mice over 7-Days Post-SAH. Sham n=8, SAH n=8-9. **A.** Individual data points are plotted with the median shown as the line plot. **B.** Same data as (**A**), but plotted with lines for each individual animal to show a tendency towards recovery. Analyzed with Scheirer-Ray-Hare test with Bonferroni post-hoc test. No significance was observed between the Sham and SAH mice at any time point.

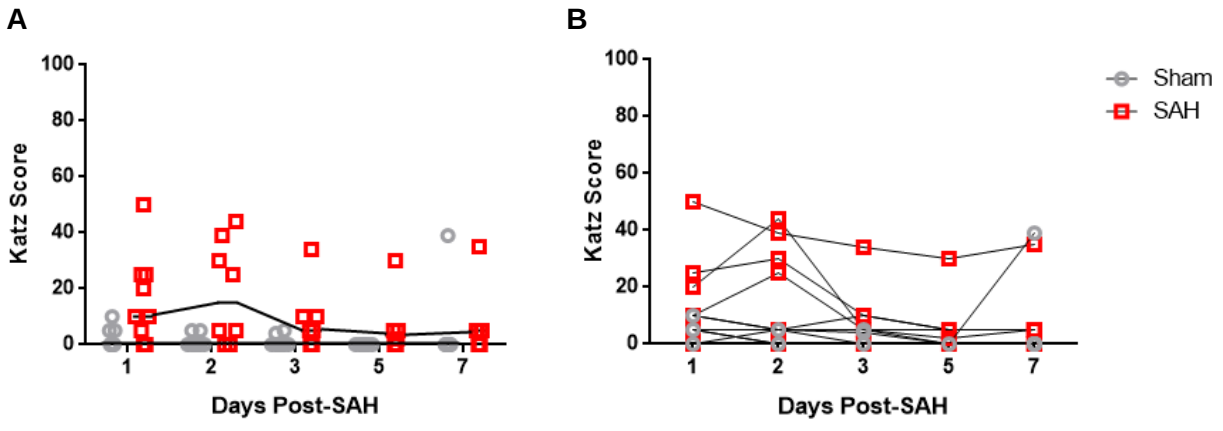

Figure S3. Functional Performance on the Katz Score for Mice over 7-Days Post-SAH. Sham n=8, SAH n=8-9. **A.** Individual data points are plotted with the median shown as the line plot. **B.** Same data as (**A**), but plotted with lines for each individual animal to show a tendency towards recovery. Analyzed with Scheirer-Ray-Hare test with Bonferroni post-hoc test. No significance was observed between the Sham and SAH mice at any time point.

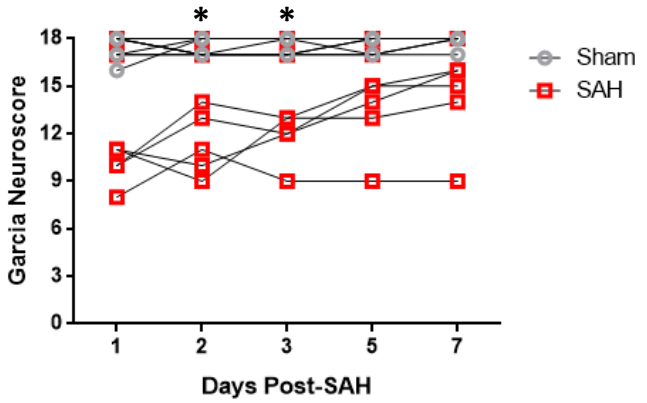

Figure S4. Neurobehavioral Performance on the Garcia Neuroscore for Mice over 7-Days Post-SAH. Individual mouse performance (reproduced from Figure 4) is connected by a line. Sham n=8, SAH n=8-9. Analyzed with Scheirer-Ray-Hare test with Bonferroni post-hoc test. * $p<0.05$ between Sham and SAH at the indicated time-point.

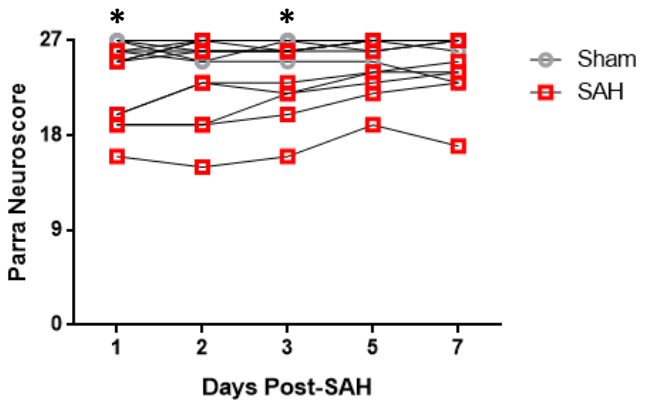

Figure S5. Functional Performance Assessed in Mice over 7-Days Post-SAH using the Parra Neuroscore. Individual mouse performance (reproduced from Figure 4) is connected by a line. Sham n=8, SAH n=8-9. Analyzed with Scheirer-Ray-Hare test with Bonferroni post-hoc test. * $p<0.05$ between Sham and SAH at the indicated time-point.

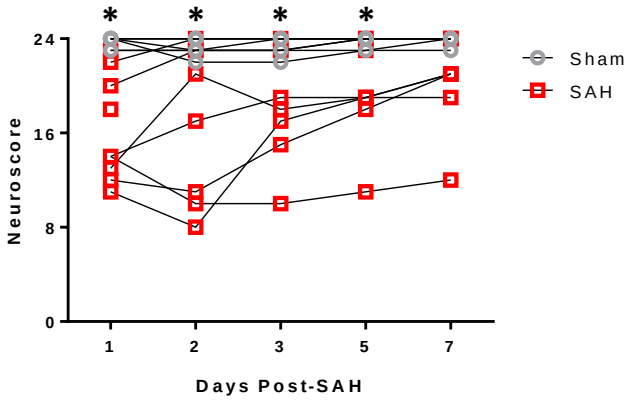

Figure S6. Neurobehavioral Performance Assessed in Mice over 7-Days Post-SAH using the Developed Composite Neuroscore. Individual mouse performance (reproduced from Figure 4) is connected by a line. Sham n=8, SAH n=8-9. Analyzed with Scheirer-Ray-Hare test with Bonferroni post-hoc test. * $p<0.05$ between Sham and SAH at the indicated time-point.

Table S1. Statistical Report for analysis of the longitudinal data in Figure 1 using Scheirer-Ray-Hare tests and a mixed effect model. Note: The Bederson Score does not meet the assumptions of the Mixed Effects Model.

	Scheirer-Ray-Hare		Mixed Effects Model	
	Reported Parameter	p-value	Reported Parameter	p-value
Bederson Score				
Groups (Sham, SAH)	16.94	<0.0001	N/A	N/A
Time (Days 1-3)	0.6651	0.7171	N/A	N/A
Interaction	0.2066	0.9019	N/A	N/A
Katz Score				
Groups (Sham, SAH)	40.12	<0.0001	-26.44	0.0011
Time (Days 1-3)	2.493	0.2875	-1.800	0.1332
Interaction	0.7559	0.6853	-0.2625	0.9180
Garcia Neuroscore				
Groups (Sham, SAH)	33.96	<0.0001	5.388	0.0004
Time (Days 1-3)	1.327	0.5151	0.2667	0.3341
Interaction	0.1998	0.9049	-0.2042	0.5481
Parra Neuroscore				
Groups (Sham, SAH)	38.95	<0.0001	5.882	0.0009
Time (Days 1-3)	2.164	0.3389	0.1667	0.2831
Interaction	0.2763	0.8710	0.1458	0.7429

Table S2. Multiple comparisons statistical report for analysis of the longitudinal data in Figure 1. Comparisons between Sham and SAH were made for each time point using Mann-Whitney U tests and corrected for multiple comparisons using a Bonferroni correction. The p-value from the Mann-Whitney U tests were compared to the Bonferroni adjusted p-value (for 3 comparisons, the Bonferroni adjusted p-value is 0.0167). Statistically significant comparisons are indicated by bold font.

	U-value	p-value
Bederson Score		
Day 1: Sham vs SAH	86.00	**0.0040**
Day 2: Sham vs SAH	85.00	**0.0107**
Day 3: Sham vs SAH	80.00	0.0177
Katz Score		
Day 1: Sham vs SAH	39.00	**<0.0001**
Day 2: Sham vs SAH	29.00	**<0.0001**
Day 3: Sham vs SAH	38.00	**0.0002**
Garcia Neuroscore		
Day 1: Sham vs SAH	36.00	**<0.0001**
Day 2: Sham vs SAH	37.50	**0.0001**
Day 3: Sham vs SAH	49.50	**0.0033**
Parra Neuroscore		
Day 1: Sham vs SAH	17.50	**<0.0001**
Day 2: Sham vs SAH	39.50	**0.0002**
Day 3: Sham vs SAH	31.00	**0.0001**

Table S3. Test Statistics and p-value for the analysis of the data in Figure 2. The data for the individual sub-tests were analyzed using Mann-Whitney U Tests. The Day 1 data from the 3-day study was used for analysis. Bold fond indicates p<0.05.

	U-Value	p-Value
Spontaneous Activity	61.50	**0.0002**
Forepaw Outstretching	110.5	**0.0057**
Climbing	59.50	**<0.0001**
Balance	43.50	**<0.0001**
Lateral Turning	95.00	**0.0044**
Walking	91.00	**0.0017**
Beam Walking	68.50	**0.0003**
Ptosis	84.00	**0.0047**
Dyspnea	128.5	**0.0417**
Facial Weakness	87.00	**0.0036**
Side Stroking	59.50	**<0.0001**
Vibrissae Touch	62.50	**<0.0001**
Visual Reflex	59.00	**0.0001**
Olfactory Reflex	83.50	**0.0008**
Tactile Reflex	136.00	0.0529
Postural Reflex	153.0	0.1930
Sound Reflex	147.0	0.1965
Righting Reflex	75.00	**0.0009**
Limb Extension	85.50	**0.0012**
Forelimb Use	52.00	**<0.0001**
Hindlimb Use	50.50	**<0.0001**

Table S4. Statistical Report for analysis of the longitudinal data in Figure 3 using Scheirer-Ray-Hare test and a mixed effect model (3-Day Study).

	Scheirer-Ray-Hare		Mixed Effects Model	
	Reported Parameter	p-value	Reported Parameter	p-value
Groups (Sham, SAH)	40.12	<0.0001	8.983	<0.0001
Time (Days 1-3)	1.917	0.3834	0.5000	0.1353
Interaction	0.4014	0.8181	-0.1875	0.7281

Table S5. Multiple comparisons statistical report for analysis of the longitudinal data in Figure 3. Comparisons between Sham and SAH were made for each time point using Mann-Whitney U tests and corrected for multiple comparisons using a Bonferroni correction. The p-value from the Mann-Whitney U tests were compared to the Bonferroni adjusted p-value (for 3 comparisons, the Bonferroni adjusted p-value is 0.0167). Statistically significant comparisons are indicated by bold font.

	U-value	p-value
Day 1: Sham vs SAH	1.500	**<0.0001**
Day 2: Sham vs SAH	34.00	**<0.0001**
Day 3: Sham vs SAH	43.00	**0.0011**

Table S6. Comparison of the ROC Curves for the Garcia Neuroscore, Parra Neuroscore, and the Developed Composite Neuroscore. AUC: area under the curve. SE: standard error of the area under the curve (from ROC curve analysis). r_{Sham}: correlation between the two test for the sham data (calculated by GraphPad Prism). r_{SAH}: correlation between the two test for the SAH data (calculated by GraphPad Prism). *r:* estimated correlation between the two tests (obtained from the table provided by Hanley and McNeil[1]).

	AUC	SE	r_{Sham}	r_{SAH}	*r*
Parra Neuroscore vs Garcia Neuroscore	0.9266 0.8875	0.04232 0.05306	0.946	0.769	0.807
Our Composite Neuroscore vs Garcia Neuroscore	0.9953 0.8875	0.00702 0.05306	0.87	0.936	0.78
Our Composite Neuroscore vs Parra Neuroscore	0.9953 0.9266	0.00702 0.04232	0.873	0.943	0.84

Table S7. Statistical Report for analysis of the longitudinal data in Figure 4 using Scheirer-Ray-Hare tests and a mixed effect model. Note: The Bederson Score does not meet the assumptions of the Mixed Effects Model.

	Scheirer-Ray-Hare		Mixed Effects Model	
	Reported Parameter	p-value	Reported Parameter	p-value
Bederson Score				
Groups (Sham, SAH)	4.155	0.0415	N/A	N/A
Time (Days 1-3, 5, 7)	3.606	0.4620	N/A	N/A
Interaction	3.606	0.4620	N/A	N/A
Katz Score				
Groups (Sham, SAH)	19.40	<0.0001	-16.53	0.0024
Time (Days 1-3, 5, 7)	3.987	0.4078	-1.802	0.0395
Interaction	0.4966	0.9738	2.121	0.0038
Garcia Neuroscore				
Groups (Sham, SAH)	29.80	<0.0001	4.892	0.0002
Time (Days 1-3, 5, 7)	2.101	0.7172	0.4278	<0.0001
Interaction	0.3057	0.9894	-0.3728	0.0006
Parra Neuroscore				
Groups (Sham, SAH)	30.93	<0.0001	5.297	0.0003
Time (Days 1-3, 5, 7)	2.306	0.6797	0.4871	<0.0001
Interaction	1.303	0.8609	-0.4784	<0.0001
Our Composite Neuroscore				
Groups (Sham, SAH)	37.99	<0.0001	7.903	<0.0001
Time (Days 1-3, 5, 7)	2.646	0.6186	0.7554	<0.0001
Interaction	1.707	0.7895	-0.7231	<0.0001

Table S8. Multiple comparisons statistical report for analysis of the longitudinal data in Figure 4. Comparisons between Sham and SAH were made for each time point using Mann-Whitney U tests and corrected for multiple comparisons using a Bonferroni correction. The p-value from the Mann-Whitney U tests were compared to the Bonferroni adjusted p-value (0.0167). Statistically significant comparisons are indicated by bold font.

	Garcia Neuroscore		**Parra Neuroscore**		**Our Composite Neuroscore**	
	U-value	p-value	U-value	p-value	U-value	p-value
Day 1: Sham vs SAH	15.00	0.0350	7.000	**0.0026**	0.0500	**0.0001**
Day 2: Sham vs SAH	8.500	**0.0132**	10.50	0.0202	9.500	**0.0143**
Day 3: Sham vs SAH	7.000	**0.0078**	4.500	**0.0016**	8.00	**0.0079**
Day 5: Sham vs SAH	11.00	0.0219	11.50	0.0232	9.500	**0.0126**
Day 7: Sham vs SAH	13.50	0.0256	18.50	0.1319	13.50	0.0256

Inter-Operator Variability

Three individuals performed the neurobehavioral assessment for the 3-day study (*i.e.* Bederson Score, Katz Score, Garcia Neuroscore, Parra Neuroscore, and the individual sub-tests). A weighted kappa statistic was used to compare the inter-operator variation between each combination of the three individuals for all tests (*i.e.* tester 1 vs tester 2, tester 1 vs tester 3, tester 2 vs tester 3).

Table S9. Inter-Operator Variation. Data from day 2 of the 3-day study is used.

	Tester 1 vs Tester 2	**Tester 1 vs Tester 3**	**Tester 2 vs Tester 3**
Scoring System	κ (Standard Error)	κ (Standard Error)	κ (Standard Error)
Bederson Score	0.9339 (0.03986)	0.9270 (0.04827)	0.9462 (0.04277)
Katz Score	0.8048 (0.08566)	0.7949 (0.10427)	0.7512 (0.11804)
Garcia Neuroscore	0.8927 (0.05197)	0.8816 (0.06764)	0.9332 (0.02685)
Parra Neuroscore	0.8802 (0.07120)	0.9192 (0.04889)	0.7980 (0.10727)
Our Composite Neuroscore	0.8920 (0.07845)	0.9135 (0.04710)	0.9395 (0.02747)

Table S10. Inter-Operator Variation. Data from day 3 of the 3-day study is used.

	Tester 1 vs Tester 2	**Tester 1 vs Tester 3**	**Tester 2 vs Tester 3**
Scoring System	κ (Standard Error)	κ (Standard Error)	κ (Standard Error)
Bederson Score	0.8235 (0.07324)	0.6087 (0.33951)	0.6250 (0.15742)
Katz Score	0.8324 (0.14991)	0.7813 (0.13133)	0.8967 (0.04568)
Garcia Neuroscore	0.9176 (0.02893)	0.9495 (0.01883)	0.9722 (0.00934)
Parra Neuroscore	0.8884 (0.07578)	0.7949 (0.11725)	0.9091 (0.05938)
Our Composite Neuroscore	0.8320 (0.09549)	0.7379 (0.14786)	0.9078 (0.04447)

Table S11. Raw Data for the Sham Mice from the 3-Day Study. Individual values represent individual mice.

Bederson Score			Katz Score			Garcia Neuroscore			Parra Neuroscore			Our Composite Neuroscore		
Day 1	Day 2	Day 3	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3
5	5	5	20	0	5	16	16	16	24	24	27	22	22	22
5	5	5	0	0	0	18	18	18	27	27	26	24	24	24
4	5	5	25	4	4	15	16	17	24	24	27	21	22	23
5	5	5	0	0	0	15	16	16	24	24	27	20	23	24
5	5	5	0	0	0	18	17	17	27	26	26	24	24	24
5	5	5	5	0	0	16	17	18	24	27	27	22	18	24
5	4	5	0	5	0	17	13	12	27	23	27	24	24	24
5	5	5	5	0	0	16	17	17	24	24	25	22	17	19
5	5	5	0	0	0	17	17	17	24	24	24	22	23	24
5	5	5	0	0	0	16	16	17	26	26	27	23	23	23
5	5	5	0	4	4	18	18	18	26	26	26	23	23	23
5	5	5	0	0	0	17	18	18	26	26	26	23	24	24
5	5	5	13	0	0	18	18	18	27	27	26	24	24	24
5	5	5	4	0	0	18	18	18	26	26	27	23	24	24
5	5	5	5	0	0	18	18	18	27	27	26	24	24	24
5	5	5	2	0	0	18	18	18	26	26	25	24	23	23

Table S12. Raw Data for the SAH Mice from the 3-Day Study. Individual values represent individual mice. Animals with missing values are those animals which died before completing the study.

Bederson Score			Katz Score			Garcia Neuroscore			Parra Neuroscore			Our Composite Neuroscore		
Day 1	Day 2	Day 3	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3
5	5	5	20	0	0	14	17	17	23	24	24	20	20	23
5	5	5	45	45	24	16	16	17	23	24	24	20	19	21
5	5	5	5	20	0	15	16	17	25	25	25	19	23	24
1	0	1	68	69	68	4	1	3	9	25	27	5	11	13
1	3	4	10	35	24	10	8	12	18	9	14	10	0	4
2	0	1	28	77	36	6	1	3	13	13	10	8	7	3
5	5	5	54	30	12	12	11	14	19	7	4	12	3	0
5	5	5	5	20	20	17	15	16	26	19	24	17	21	24
5	5	5	5	0	0	15	18	18	24	27	27	19	24	24
3	5	5	50	39	34	8	11	9	15	15	18	14	10	10
5	5	5	25	13	9	17	15	17	23	23	25	20	10	10
5	0	2	17	5	58	14	4	7	21	22	24	13	23	23
2	5	5	69	74	14	8	16	11	15	11	15	9	17	21
4	2	2	17	0	0	11	10	7	21	25	22	13	7	9
0	5	5	100	91	71	3	17	18	8	18	16	3	20	12
4	0		31	96		11	1		19	8		10	1	
5	0		5	45		17	10		15	12		18	10	
0			86			5			12			3		
5			0			17			23			18		
0			100			3			8			3		

Table S13. Raw Data for the Sham Mice from the 7-Day Study for Bederson Score and Katz Score. Individual values represent individual mice.

Bederson Score					Katz Score				
Day 1	Day 2	Day 3	Day 5	Day 7	Day 1	Day 2	Day 3	Day 5	Day 7
5	5	5	5	5	0	5	5	0	39
5	5	5	5	5	0	0	0	0	0
5	5	5	5	5	0	0	0	0	0
5	5	5	5	5	0	0	0	0	0
5	5	5	5	5	0	0	0	0	0
5	5	5	5	5	5	0	0	0	0
5	5	5	5	5	5	0	0	0	0
5	5	5	5	5	10	5	4	0	0

Table S14. Raw Data for the Sham Mice from the 7-Day Study for the Garcia Neuroscore, Parra Neuroscore, and Our Composite Neuroscore. Individual values represent individual mice.

Garcia Neuroscore					Parra Neuroscore					Our Composite Neuroscore				
Day 1	Day 2	Day 3	Day 5	Day 7	Day 1	Day 2	Day 3	Day 5	Day 7	Day 1	Day 2	Day 3	Day 5	Day 7
18	17	17	17	17	27	25	25	25	23	24	22	22	23	23
18	18	18	18	18	27	27	27	27	27	24	24	24	24	24
18	18	18	18	18	27	27	27	27	27	24	24	24	24	24
18	18	18	18	18	27	27	27	27	27	24	24	24	24	24
18	17	17	18	18	27	26	26	27	26	24	23	23	24	24
17	17	18	17	18	26	25	27	26	27	24	24	24	24	24
17	18	18	18	18	25	27	27	27	27	24	24	24	24	24
16	18	18	18	18	25	27	27	27	27	23	23	24	24	24

Table S15. Raw Data for the SAH Mice from the 7-Day Study for Bederson Score and Katz Score. Individual values represent individual mice. Animals with missing values are those animals which died before completing the study.

Bederson Score					Katz Score				
Day 1	Day 2	Day 3	Day 5	Day 7	Day 1	Day 2	Day 3	Day 5	Day 7
3	5	5	5	2	50	39	34	30	35
5	5	5	5	5	5	5	0	0	0
5	5	5	5	5	20	44	5	2	5
5	5	5	5	5	0	0	0	0	0
4	5	4	5	5	10	25	5	5	5
5	5	5	5	5	0	0	0	0	0
5	5	5	5	5	25	30	10	5	5
5	5	5	5	5	10	5	10	5	5
5	5				25				

Table S16. Raw Data for the SAH Mice from the 7-Day Study for the Garcia Neuroscore, Parra Neuroscore, and Our Composite Neuroscore. Individual values represent individual mice. Animals with missing values are those animals which died before completing the study.

Garcia Neuroscore					Parra Neuroscore					Our Composite Neuroscore				
Day 1	Day 2	Day 3	Day 5	Day 7	Day 1	Day 2	Day 3	Day 5	Day 7	Day 1	Day 2	Day 3	Day 5	Day 7
8	11	9	9	9	16	15	16	19	17	17	10	10	11	12
17	17	17	17	18	25	26	26	26	27	20	23	23	23	24
11	9	13	15	15	19	19	22	23	24	11	8	17	19	19
18	17	17	18	18	26	26	26	27	27	23	23	23	24	24
10	13	12	15	16	20	23	23	24	25	14	17	19	19	21
18	18	18	18	18	26	27	26	27	27	22	24	24	24	24
11	10	12	14	16	19	19	20	22	23	12	11	15	18	21
10	14	13	13	14	20	23	22	24	24	13	21	18	19	21
17					25					18				

Supplemental Reference:

1. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983;148:839-843