# Inter-rater agreement for consensus definitions of delayed ischemic events following aneurysmal subarachnoid hemorrhage

**Sahar Zafar, MD**[1], **M. Brandon Westover, MD, PhD**[1], **Nicolas Gaspard, MD, PhD**[2], **Emily Gilmore, MD**[3], **Brandon Foreman, MD**[4], **Kathryn O'Connor, BS**[1], and **Eric S. Rosenthal, MD**[1]

[1]Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

[2]Department of Neurology, Université Libre de Bruxelles, Brussels, Belgium

[3]Department of Neurology, Yale University School of Medicine, New Haven, CT, USA

[4]Department of Neurology and Rehabilitation Medicine, UC College of Medicine, Cincinnati, OH, USA

## Abstract

**Background**—Thirty percent of subarachnoid hemorrhage (SAH) patients experience delayed cerebral ischemia (DCI) or delayed ischemic neurological decline (DIND). Variability in the definitions of delayed ischemia make outcome studies difficult to compare. A recent consensus statement advocates standardized definitions for delayed ischemia in clinical trials of SAH. We sought to evaluate the inter-rater agreement (IRA) of these definitions.

**Methods**—Based on consensus definitions, we assessed for: 1. Delayed cerebral infarction (DCIN), defined as radiographic cerebral infarction; 2. DIND Type 1 (DIND1), defined as focal neurological decline; and 3. DIND2, defined as a global decline in arousal. Five neurologists retrospectively reviewed electronic records of 58 SAH patients. Three reviewers had access to and reviewed neuroradiology imaging. We assessed IRA using Gwet's kappa statistic.

**Results**—IRA statistics were excellent (95.83%) for overall agreement on the presence or absence of any delayed ischemic event (DIND1, DIND2 or DCIN). Agreement was "moderate" for specifically identifying DIND1 (56.58%) and DIND2 (48.66%) events. We observed greater agreement for DIND1 when there was a significant focal motor decline of at least 1 point in the motor score. There was fair agreement (39.20%) for identifying DCIN; CT imaging was the predominant modality.

**Conclusion**—Consensus definitions for delayed cerebral ischemia yielded near-perfect overall agreement and can thus be applied in future large scale studies. However, a strict process of adjudication, explicit thresholds for determining focal neurologic decline and MRI techniques that

*Corresponding Author: Eric S. Rosenthal, 55 Fruit Street, Massachusetts General Hospital, Boston, MA, USA 02114, Tel: 617-726-3311, Fax: 617-726-9250, ; Email: erosenthal@mgh.harvard.edu

better discriminate edema from infarction appear critical for reproducibility of determination of specific outcome phenotypes, and will be important for successful clinical trials.

### Keywords

subarachnoid hemorrhage; ischemia; vasospasm; critical care; nomenclature

## Introduction

Aneurysmal subarachnoid hemorrhage (SAH) is a neurological emergency with an incidence of 2-22 per 100,000 per year (Van Gijn and Rinkel, 2001). Among survivors, approximately one third to one half suffer from long term neurologic disability (Suarez et al., 2006; Al-Tamimi et al., 2010). Delayed cerebral ischemia (DCI) is one of the most significant complications in the days that ensue after an acute aneurysmal SAH, occurring in up to 30% of patients (Roos et al., 2000). DCI typically occurs 4 to 12 days after the initial hemorrhage and is a major cause of morbidity and death among patients (Hijdra et al., 1986; Roos et al., 2000). Much research has been dedicated to identification of predictors and early detection of delayed cerebral ischemia but ascertainability of these events has received scant attention.

DCI has been defined for research purposes using variations on three basic criteria: (a) delayed focal neurological deficits, (b) delayed decline in level of consciousness, or (c) cerebral infarction found on imaging. All three definitions usually include the proviso that the ischemic event should neither be the result of aneurysm treatment nor explained by other factors such as edema, hydrocephalus or infection (Washington et al., 2011).

Studies of DCI have employed various terms for the phenomenon, including "vasospasm", "delayed ischemic neurological deficit", "delayed cerebral infarction", "delayed ischemic deficit", "delayed neurological deficit", "secondary cerebral ischemia", "clinical vasospasm", "symptomatic vasospasm", "symptomatic ischemia" and "cerebral infarction" (Vergouwen et al., 2010a). Distinction needs to be made between vasospasm, which is angiographic evidence of vessel narrowing, and delayed clinical neurological deficits, particularly because the two do not always co-exist (Washington et al., 2011). Moreover, there is considerable variation in the use of clinical events and radiographic infarction to define delayed cerebral ischemia (Vergouwen et al., 2010a).

Given the variation in definitions, a recent consensus statement advocated standardized definitions for Delayed Cerebral Ischemia and Delayed Ischemic Neurological Deficits in clinical trials of SAH(Vergouwen et al., 2010a). We sought to operationalize these definitions by making them more explicit, and to evaluate the inter-rater agreement (IRA) of these definitions, to determine the extent to which they can be applied consistently in the context of large-scale observational research.

## Methods

Based on published consensus definitions (Vergouwen et al., 2010a), we categorized delayed cerebral ischemia into three event types:

- Delayed cerebral infarction (DCIN), defined as radiographic cerebral infarction

- Delayed ischemic neurological decline Type 1 (DIND1), defined as focal neurological decline; and

- Delayed ischemic neurologic decline Type 2 (DIND2), defined as global decline in arousal.

Our definitions for DIND and DCIN, with explicit exclusion criteria are listed in Table 1. Under an investigational review board approved protocol, five neurologists from two different institutions retrospectively reviewed detailed electronic records of 58 aneurysmal SAH patients who had undergone advanced physiological monitoring at a single institution from September 2011 to January 2014. All reviewers were board-certified neurologists; three were neurocritical care physicians; and two were neurologists with expertise in critical care neuromonitoring and clinical neurophysiology. Three reviewers were from the same institution where the patients underwent monitoring. All reviewers had access to an identical standardized version of each patient's medical record documentation of the admission presentation, clinical event notes, neurological examinations, laboratory results, transcranial doppler ultrasound findings, and imaging reports. Three reviewers additionally had access to Digital Imaging and Communications in Medicine (DICOM) based neuroimaging data (AMICAS, Inc., Waltham, MA). Each reviewer individually went through a standardized version of daily patient records, using the 24-hour events, clinical exam, radiographic and laboratory data subsections from each record to identify whether or not a qualifying DCIN or DIND event occurred on each hospital day.

Inter-rater agreement was calculated using Gwet's multi-rater agreement coefficient AC1 (Gwet, 2008a; Gwet, 2010b). In assessments of inter-rater agreement, a portion of the observed percent agreement (PA) is assumed to be attributable to chance (PC), and inter-rater agreement statistical methods are used to estimate the percent agreement beyond chance or kappa ($\kappa$).

$$\kappa = \frac{PA - PC}{100 - PC}$$

Several statistical kappa measures are available for evaluating inter-rater agreement. These include Cohen's kappa, used for 2 raters; Fleiss' multi-rater kappa; and Gwet's multi-rater kappa. We chose Gwet's multi-rater agreement coefficients AC1 for our analysis over Cohen's and Fleiss' kappa statistics, because the latter measures may be less accurate when there is a high degree of agreement or disagreement (Feinstein and Cicchetti, 1990; Gwet, 2008a). We categorized $\kappa$ values following the convention: slight agreement 10-20%; fair agreement 20-40%; moderate agreement 40-60%; substantial agreement 60-80%; near perfect agreement 80-100% (Landis and Koch, 1977). We also calculated 95% confidence intervals for the estimated $\kappa$ values using the Jackknife method (Tukey, 1958; Gwet, 2010b). Statistical calculations and figure creation were performed using the Matlab Statistics Toolbox and custom software developed in-house using MATLAB (The Mathworks, Natick, MA).

## Results

The main demographic and clinical characteristics are shown in Table 2. The majority of our patients had Hunt and Hess (HH) 3 and Fisher grade 3 subarachnoid hemorrhages, with distal ICA aneurysms being the most common source of hemorrhage.

The percentages of observed agreement and estimated chance-corrected level of inter-rater agreement (kappa value, $\kappa$), with 95% confidence intervals, are shown in Figure 1a and 1b. All between-rater pairwise percent agreement values are shown using confusion matrices in Figure 2.

There was near perfect overall agreement ($\kappa$=95.83% CI=95.59-96.06) regarding the presence or absence of any of DCIN, DIND1 or DIND2 events without regard to subtype. We found moderate agreement specifically on DIND1 events ($\kappa$=56.58%) and DIND2 events ($\kappa$=48.66%) when restricting the analysis only to events identified by at least one rater. There was fair agreement when restricting the analysis to DCIN events identified by at least one reader, with a percent agreement of 46.67% and $\kappa$ of 39.20%. There was a higher degree of agreement between the three reviewers from the same institution; kappa values were 72.71% and 69.58 % for DIND1 and DIND2 events respectively.

We reviewed in detail the clinical features of those DIND1 events with unanimous agreement and those with poor agreement (Table 3 and 4). We found greater agreement when there was at least a 1-point decline in motor strength testing using the Medical Research Council (MRC) scale (Medical Research Council, 1976). There appeared to be less agreement with subtle changes on exam (Table 4). This difficulty discriminating subtle changes was seen more often when the baseline strength on the motor exam was zero, in which case the change in exam was described as a change from localization or withdrawal to a flexion or extension posturing response.

## Discussion

20-40 % of patients with SAH die or have long term neurological disability, often related to delayed cerebral ischemia (Washington et al., 2011). Accurate detection of DCI provides an opportunity to intervene and improve outcomes. Although clinical detection of DCI on neurological examination has been considered problematic and inferior to radiologic outcomes, we found excellent overall agreement using consensus definitions when not constrained by subtype. Even when subtyped, clinical episodes of DIND were more reliably ascertained than radiologic DCIN; moderate agreement was observed for a focal neurologic decline and for a global decline in arousal, but only fair agreement was observed for radiologic DCIN. These results provide encouragement that clinicians viewing identical clinical data can agree on clinical outcomes for the purpose of large-scale observational studies. However, these findings are in contrast to consensus recommendations to utilize radiologic DCIN as the preferred and most reliable method of ascertaining events of delayed ischemic events (Vergouwen et al., 2010a).

Identifying focal neurological events in our cohort however did not have perfect inter-rater agreement. One challenge may have been the variety of patient presentations, including but

not limited to focal weakness, aphasia, mutism, and visual deficits. There were also apparent differences in thresholds among physicians for diagnosing a focal neurological decline. There appeared to be a greater degree of agreement when there was a significant change in the motor exam, particularly when the baseline motor strength was 3 or greater. Agreement was less for subtle changes in neurologic exam, and for the appearance of new cranial nerve deficits or changes in the language exam.

Moderate agreement for global declines in arousal could have also resulted from different thresholds among physicians for calling an event a global decline. Half the patients were intubated and sedated, which can confound detection of exam changes, leading to decreased agreement. The application of clinical criteria to define delayed ischemic neurologic decline is also challenging in patients with poor baseline exams, and cases of delayed ischemia may be missed (Schmidt et al., 2008; Vergouwen et al., 2010).

Overall agreement on DIND events could also be affected by the exclusion criteria we established. We excluded patients who were suspected to have seizures, infection, toxic and metabolic encephalopathy, sedatives and hydrocephalus as possible underlying etiologies for focal and global changes in their neurological exam. However, applying these complex exclusion criteria involves clinical judgment and thereby creates further opportunity for subjective disagreement. Beyond the issue of agreement, it is possible that patients could have simultaneous delayed cerebral ischemia with co-morbid occurrence of infections or other exclusion events, and it may be more appropriate to employ definitions in which DIND or DCIN is not a diagnosis of exclusion.

Surprisingly, we found only fair agreement for radiographic DCIN events. This discrepancy is likely due to the predominant use of non-contrast CT scans in our cohort. Non-contrast CT scans have limitations in detecting early cerebral ischemia (Dankbaar et al., 2009), including early infarction at the time of SAH, and in differentiating edema from infarction. The use of MRI to distinguish edema from ischemia, and thereby better determine DCIN, would likely be more accurate. Asymptomatic radiographic infarction can occur in at least 4 % of SAH patients (Schmidt et al., 2008; Vergouwen et al.,2011b) and is highly associated with baseline coma and with worse outcomes. This highlights the importance of obtaining both baseline and follow-up imaging in patients of poor initial clinical grade. A 3-tiered combined clinical and imaging reference standard that uses clinical exam, TCD, CTA, MRA and DSA data has also been proposed to increase consistency in the definition of delayed cerebral ischemia, and to serve as a tool for outcome studies (Sanelli et al., 2014). However, this approach would require serial imaging to discriminate DCIN from peri-procedure lesions, particularly in comatose patients

A limitation of our study is its retrospective nature, as DIND events are challenging to determine not only retrospectively, but also in real-time. Although all reviewers had access to daily progress notes and clinical data, only three reviewers had access to imaging data, hence the inter-rater agreement for imaging data is limited to the three reviewers within the same institution. Of note, we found higher inter-rater agreement for the three reviewers at the same institution where patients were admitted, which may reflect an institutional approach to applying the definitions used in the study.

## Conclusion

In conclusion, we found excellent overall agreement using consensus definitions of delayed ischemic events. However, there are challenges to determining DCI as outlined by our study, and a strict process of adjudication, and both an explicit threshold for determining focal neurologic decline and use of MRI imaging to distinguish edema from infarction may further improve validity of DCI diagnoses. Combining more reliable clinical criteria with ancillary data from better quality imaging data, such as perfusion imaging, MRI, and transcranial Doppler, may provide more robust outcome measures for clinical trials and quality improvement studies. Microvascular, electrophysiological, and metabolic etiologies have been proposed as additional mechanisms causing delayed cerebral ischemia in the absence of proximal vasospasm or radiologic infarction (Westermaier et al., 2014). Given this incomplete overlap, a severe and persistent deterioration in neuromonitoring trends from any cause may also warrant consideration as an additional hospital outcome measure deserving validation against long-term clinical outcomes.

## Acknowledgments

## References

Al-Tamimi YZ, Orsi NM, Quinn AC, Homer-Vanniasinkam S, Ross SA. A review of delayed ischemic neurologic deficit following aneurysmal subarachnoid hemorrhage: historical overview, current treatment, and pathophysiology. World Neurosurg. 2010; 73:654–67. [PubMed: 20934153]

Dankbaar JW, de Rooij NK, Velthuis BK, Frijns CJM, Rinkel GJE, van der Schaaf IC. Diagnosing delayed cerebral ischemia with different CT modalities in patients with subarachnoid hemorrhage with clinical deterioration. Stroke J Cereb Circ. 2009; 40:3493–8.

Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990; 43:543–9. [PubMed: 2348207]

Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol. 2008a; 61:29–48. [PubMed: 18482474]

Gwet, KL. Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters. Gaithersburg, MD: Advanced Analytics, LLC; 2010b.

Hijdra A, Van Gijn J, Stefanko S, Van Dongen KJ, Vermeulen M, Van Crevel H. Delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage: clinicoanatomic correlations. Neurology. 1986; 36:329–33. [PubMed: 3951698]

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159–74. [PubMed: 843571]

Medical Research Council. Medical Research Council scale. Aids to examination of the peripheral nervous system. Memorandum no 45. 1976

Roos YB, de Haan RJ, Beenen LF, Groen RJ, Albrecht KW, Vermeulen M. Complications and outcome in patients with aneurysmal subarachnoid haemorrhage: a prospective hospital based cohort study in the Netherlands. J Neurol Neurosurg Psychiatry. 2000; 68:337–41. [PubMed: 10675216]

Sanelli PC, Kishore S, Gupta A, Mangat H, Rosengart A, Kamel H, et al. Delayed Cerebral Ischemia in Aneurysmal Subarachnoid Hemorrhage: Proposal of an Evidence-Based Combined Clinical and Imaging Reference Standard. AJNR Am J Neuroradiol. 2014; 35:2209–2214. [PubMed: 24263697]

Schmidt JM, Wartenberg KE, Fernandez A, Claassen J, Rincon F, Ostapkovich ND, et al. Frequency and clinical impact of asymptomatic cerebral infarction due to vasospasm after subarachnoid hemorrhage. J Neurosurg. 2008; 109:1052–9. [PubMed: 19035719]

Suarez JI, Tarr RW, Selman WR. Aneurysmal Subarachnoid Hemorrhage. N Engl J Med. 2006; 26:354, 387–96.

Tukey J. Bias and confidence in not-quite large samples. Annals of Mathematical Statistics. 1958; 29:614.

Van Gijn J, Rinkel GJ. Subarachnoid haemorrhage: diagnosis, causes and management. Brain J Neurol. 2001; 124:249–78.

Vergouwen MDI, Vermeulen M, van Gijn J, Rinkel GJE, Wijdicks EF, Muizelaar JP, et al. Definition of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage as an outcome event in clinical trials and observational studies: proposal of a multidisciplinary research group. Stroke J Cereb Circ. 2010a; 41:2391–5.

Vergouwen MDI. Participants in the International Multi-Disciplinary Consensus Conference on the Critical Care Management of Subarachnoid Hemorrhage. Vasospasm versus delayed cerebral ischemia as an outcome event in clinical trials and observational studies. Neurocrit Care. 2011b; 15:308–11. [PubMed: 21748502]

Washington CW, Zipfel GJ. Participants in the International Multi-disciplinary Consensus Conference on the Critical Care Management of Subarachnoid Hemorrhage. Detection and monitoring of vasospasm and delayed cerebral ischemia: a review and assessment of the literature. Neurocrit Care. 2011; 15:312–7. [PubMed: 21748499]

Westermaier T, Pham M, Stetter C, Willner N, Solymosi L, et al. Value of transcranial doppler, perfusion-CT and neurological evaluation to forecast secondary ischemia after aneurysmal SAH. Neurocrit Care. 2014 Jun; 20(3):406–12. [PubMed: 23982597]
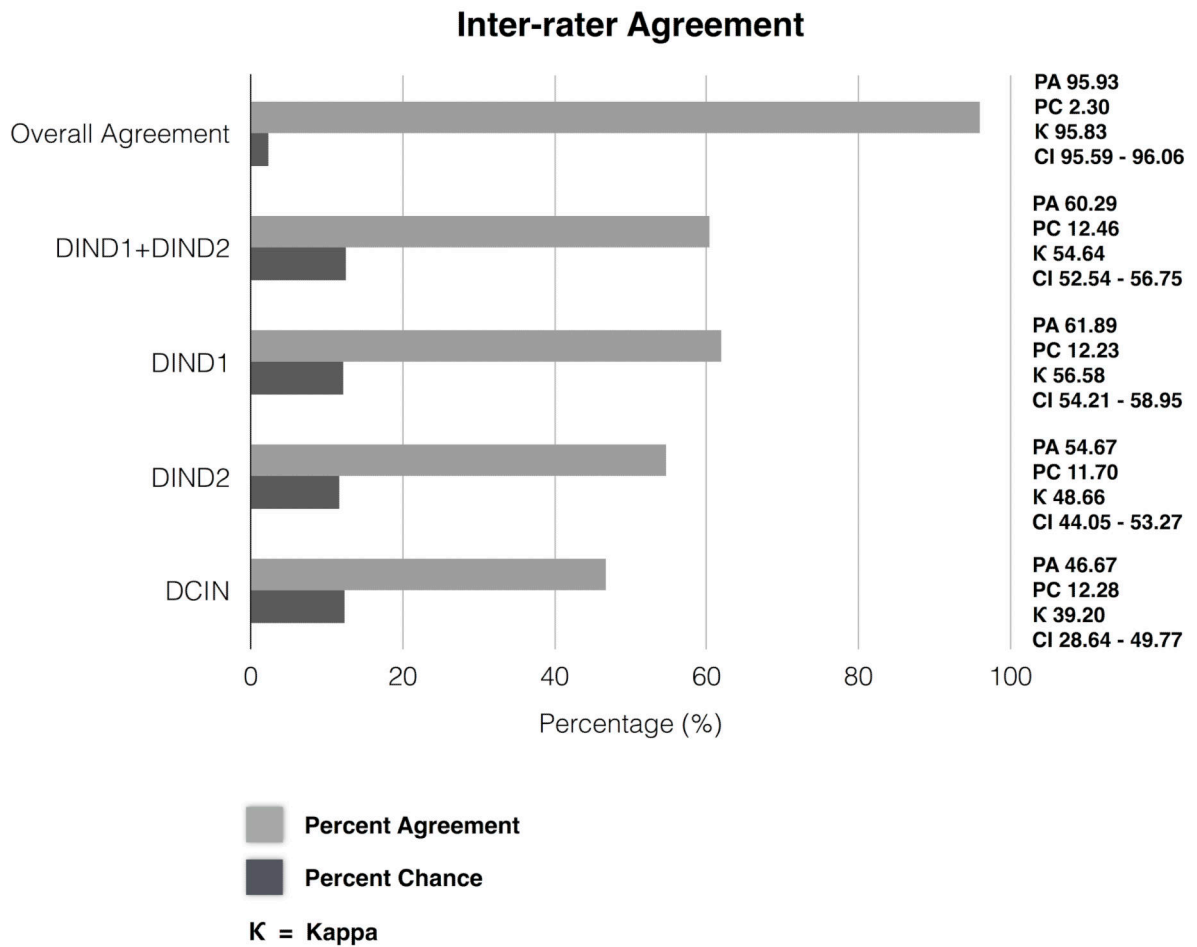
## Inter-rater Agreement



**Figure 1a. Inter-rater agreement (IRA) for all raters**

In calculating IRA, a portion of the observed percent agreement (PA), is assumed to be attributable to chance (PC), and inter-rater agreement statistical methods are used to determine the percent agreement beyond chance or kappa (κ).

$$\kappa = \frac{PA - PC}{100 - PC}$$

κ 10-20% = slight agreement; κ 20-40% =fair agreement; κ 40-60%=moderate agreement; κ 60-80%=substantial agreement; κ 80-100%= near perfect agreement. Overall PA, PC, κ and 95% confidence interval (CI) is shown, along with PA, PC, κ values and CI for each event subtype (DIND1, DIND2 and DCIN). DIND; Delayed ischemic neurological decline DCIN; Delayed cerebral infarction
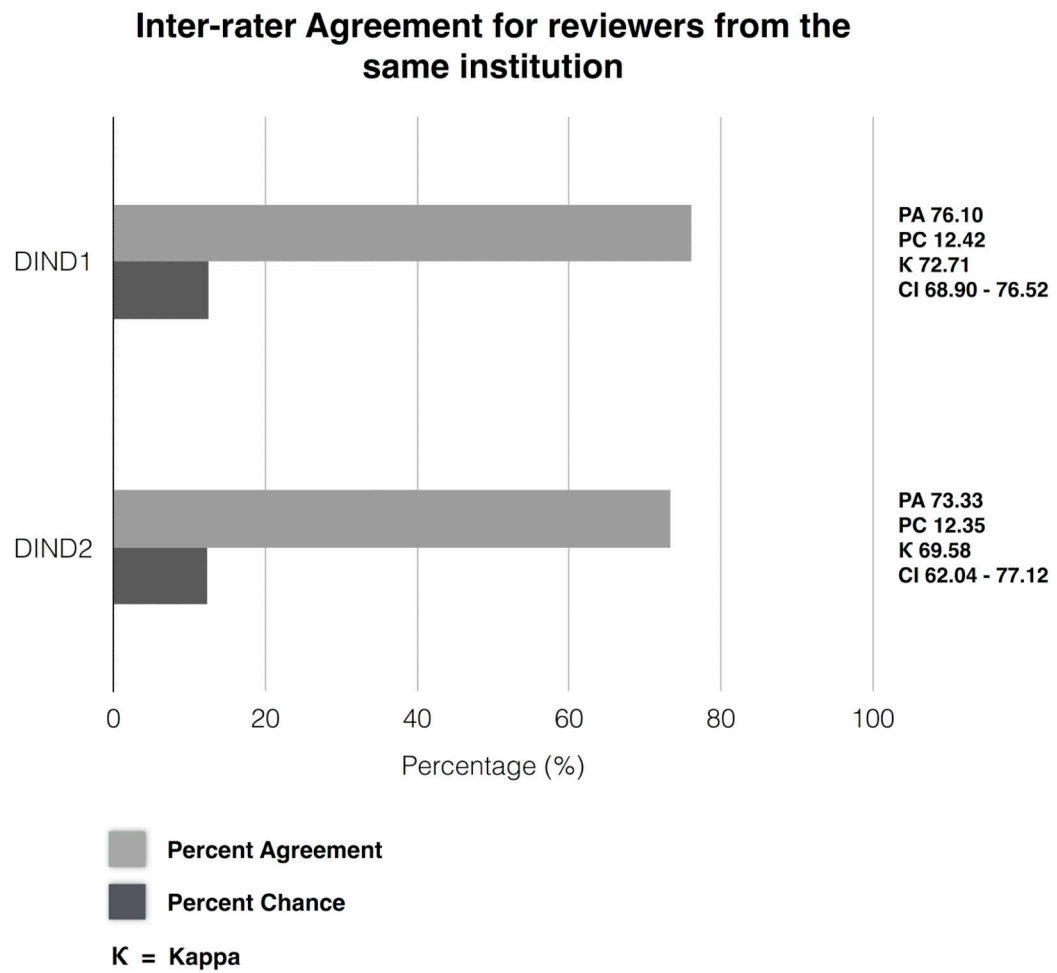
## Inter-rater Agreement for reviewers from the same institution



**DIND1**
PA 76.10
PC 12.42
K 72.71
CI 68.90 - 76.52

**DIND2**
PA 73.33
PC 12.35
K 69.58
CI 62.04 - 77.12

Percentage (%)

■ **Percent Agreement**

■ **Percent Chance**

**K = Kappa**

**Figure 1b. Inter-rater agreement for the 3 reviewers from the same institution**
PA, PC, $\kappa$ and CI are shown for DIND1 and DIND2 event subtypes for the 3 raters from the same institution where patients were admitted.
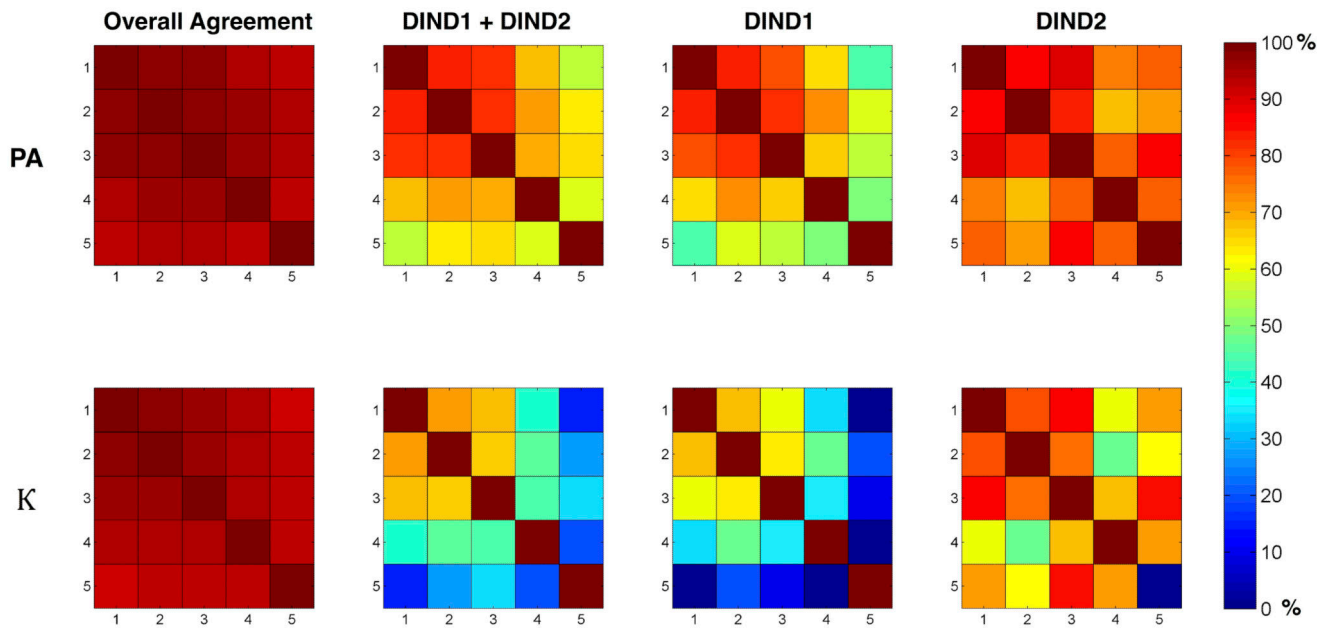
# Agreement Matrices



**Figure 2. Agreement matrices**

Pair wise percent agreement (PA) and kappa ($\kappa$) values for all 5 raters are shown. Raters 1, 2 and 3 were from one institution and raters 4 and 5 were from the second institution. Raters 1,3 and 5 were Neurocritical care physicians. DIND; Delayed ischemic neurological decline DCIN; Delayed cerebral infarction

**Table 1**
**Definition of DIND and DCIN and exclusion criteria**

| Definition |
| --- |
| Delayed Ischemic Neurological Decline (DIND): |

Type I – New focal neurological deficit lasting at least 1 hour, present on at least 2 serial assessments not attributable to an event on the list of DIND Exclusion Events.

Type II –GCS decline in the total or sub score of significant magnitude ( 2 points) and rapidity (within a 4 hour window), which is sustained (lasting at least 1 hour) and not attributable to a DIND exclusion event.

Delayed Cerebral Infarction (DCIN):
Radiologic signs of infarction satisfying one of the following:

a. Seen >48 hours after any intracranial procedure and absent on intervening imaging.

b. Seen >48 hours after any intracranial procedure without intervening imaging in a location distant from the operative site and not considered to be caused by a procedure (e.g. an embolic complication of angiography).

c. Seen <48 hours after an intracranial procedure in the setting of severe vasospasm

d. Seen <48 hours after an intracranial procedure in a vascular or watershed territory distinct from operative site, not considered to be caused by a procedure (e.g. angiography)

Exclusion criteria for DIND

1 Re-bleeding

2 ICP spike: an event at time T is excluded by ICP if, within a 4 hour window centered on time T (i.e. two hours before and after T) there are at least two ICP values >20 mmHg.

3 Acute hydrocephalus

4 Acute metabolic abnormality**

5 New evidence of infection

6 Seizure (confirmed on EEG)

7 Change in sedative medications temporally associated with clinical symptoms

8 Neurosurgical procedure within 24 hours preceding the event and considered causative

*Fever by itself, without further compelling evidence of infection, is not included on the list of DIND Exclusion Events.
** See below for specific definitions.

1 Serum Sodium:

    a. Range: <130 and >155

    b. Rate of change: Change (↑ or ↓) of 10 meq within 24 hours preceding the event

2 BUN:

    a. Range: >60

    b. Rate of change: ↑ 20 within 24 hours preceding the event

3 Glucose:

    a. Range: < 40 and >300

    b. Rate of change: ↑ 100 within 24 hours preceding the event

4 Hemoglobina.

    a. Range: <7

    b. Rate of change: ↓ 2 within 24 hours preceding the event

5 PO2:

    a. Range: <60

    b. Rate of change: ↓ 20 within 24 hours preceding the event

6 WBC:

| Definition | | |
|---|---|---|
| | a. | Range: > 20 |
| | b. | ↑ 10 within 24 hours preceding the event |

**Table 2**
**Demographics and Clinical Characteristics**

| Demographics | |
|---|---|
| Mean age (yrs) | 56.9 (STD 15.2) |
| Gender | 40 (69.0%) Female |
| | 18 (31.0%%) Male |
| **Clinical Characteristics** | |
| Hunt and Hess Grade | n (%) |
| Grade 1 | 11 (19.0%) |
| Grade 2 | 11 (19.0%) |
| Grade 3 | 14 (24.1%) |
| Grade 4 | 11 (19.0%) |
| Grade 5 | 11 (19.0%) |
| Fisher Grade | n (%) |
| Grade 1 | 0 (0.00%) |
| Grade 2 | 0 (0.00%) |
| Grade 3 | 49 (84.5%) |
| Grade 4 | 9 (15.5%) |
| Location of Aneurysm | n (%) |
| Distal ICA | 14 (24.1%) |
| ACA | 6 (10.3%) |
| MCA | 7 (12.1%) |
| PCA | 2 (3.4%) |
| Acomm | 12 (20.1%) |
| Pcomm | 8 (13.8%) |
| Basilar | 4 (6.9%) |
| Other | 5 (8.6%) |
| Total number of DIND events | 67 |
| Total number of DIND events with unanimous agreement between raters | 9/67 (13.4%) |
| Total number of DIND events where at least 3/5 raters agreed | 26/67 (38.8%) |
| Median delay from date of bleed to detection of DIND | 5 days |
| **Other clinical characteristics** | n (%) |
| Hydrocephalus requiring EVD | 45 (77.6%) |
| Seizures on admission | 9 (15.5 %) |
| Post op stroke/ICH | 4 (6.9%) |
| Intra op re-rupture | 5 (8.6%) |
| Need for mechanical ventilation | 30 (51.7%) |

Abbreviations; ICA: internal carotid artery; ACA: anterior cerebral artery; MCA: middle cerebral artery; PCA: posterior cerebral artery; Acomm: anterior communicating artery; Pcomm: posterior communicating artery; EVD: external ventricular drain.

**Table 3**
**Sample events with unanimous agreement**

| Change in motor strength exam | Comments |
|---|---|
| Right arm 4->1 Right leg 3->1 | Baseline exam; following commands, full strength in bilateral arms and antigravity in bilateral legs |
| Right arm 1->0 | Baseline exam; no eye opening. Had baseline flexion of left arm, localizing right arm and triple flexion of both legs. Right arm went from localizing to flexion posturing |
| Left leg 4->2 | Baseline exam; following commands- full strength bilateral arms and legs |
| No change in motor score, "decreased activation on the right" described in the exam | Baseline exam; opening eyes, and following commands in all four extremities |
| Left arm 1-> 0 | Baseline exam; Intermittently following commands, localizing in the right upper extremity and triple flexion of bilateral lower extremities |
| Left leg 2->0, and stopped following commands | Baseline exam; following commands, localizing bilateral upper extremities, and withdrawing bilateral lower extremities |
| Right arm 2->0 Right leg 1->0 | Baseline exam; follows commands, strength 0/5 on left arm and leg and 5/5 on right arm and leg |
| Left leg 2-> 0 Left arm 2-> 0 | Baseline exam; following commands, moving all extremities antigravity |
| 3->2 bilateral legs and abulia | Baseline exam; following commands, language fluent, full strength in all extremities |

**Table 4**

**Sample events with poor agreement**

| Description of exam change |
|---|
| Left arm brisker than right , no change in motor strength score |
| Transient left facial droop and field cut, resolved with BP elevation |
| Exam described as "weaker on left lower extremity" |
| Bilateral arms went from withdrawal to trace movements |
| Patient went from following commands to localizing in arms |
| Left facial droop in a patient with previous postoperative left arm and leg weakness |
| Left ptosis |
| Decreased movement of right foot, with no change in motor score |
| Left abducens palsy |
| Left arm and leg from 5/5 -> 3/5 |
| Increased perseveration on exam with no change in motor score |
| Decreased movement of right hand, with no change in motor score |