Agreement Regarding Diagnosis of Transient Ischemic Attack Fairly Low Among Stroke-Trained Neurologists

James Castle, MD; Michael Mlynash, MD, MS; Karming Lee, MD; Anna Finley Caulfield, MD; Connie Wolford, RN, MSN, NP-C; Stephanie Kemp, BS; Scott Hamilton, PhD; Gregory W. Albers, MD; Jean-Marc Olivot, MD, PhD

Background and Purpose—Agreement between physicians to define the likelihood of a transient ischemic attack (TIA) remains poor. Several studies have compared neurologists with nonneurologists, and neurologists among themselves, but not between fellowship-trained stroke neurologists. We investigated the diagnostic agreement in 55 patients with suspected TIA.

Methods—The history and physical examination findings of 55 patients referred to the Stanford TIA clinic from the Stanford emergency room were blindly reviewed by 3 fellowship-trained stroke neurologists who had no knowledge of any test results or patient outcomes. Each patient’s presentation was rated as to the likelihood that the presentation was consistent with TIA. We used 3 different scales (2-, 3-, and 4-point scales) to define TIA likelihood. We assessed global agreement between the raters and evaluated the biases related to individual raters and scale type.

Results—The agreement between fellowship-trained stroke neurologists remained poor regardless of the rating system used and the statistical test used to measure it. Difference in rating bias among all raters was significant for each scale: \( P = 0.001, 0.012, \) and \( < 0.001 \). In addition, for each reviewer, the rate of labeling an event an “unlikely TIA” progressively decreased with the number of points that composed the scale.

Conclusions—TIA remains a highly subjective diagnosis, even among stroke subspecialists. The use of confirmatory testing beyond clinical judgment is needed to help solidify the diagnosis. Caution should be used when diagnosing an event as a possible TIA. (Stroke. 2010;41:1367-1370.)

Key Words: diagnosis ■ TIA ■ interrater agreement

Transient ischemic attack (TIA) is diagnosed among 300 000 patients every year in the United States.\(^1\,^2\) A recent meta-analysis shows that the risk of stroke after a TIA is 5.2% (95% CI, 3.9 to 6.5)\(^3\) during the first week and that most of the strokes occur within the first 72 hours, enforcing the notion that TIA is a vascular emergency. TIA management requires an emergent comprehensive workup, including brain and arterial imaging, cardiac evaluation, and risk factor assessment,\(^4\) all essential to tailor an efficient stroke prevention strategy. Hence, the clinical diagnosis of a neurological symptom of vascular origin that will launch this evaluation is critical.

Several studies conducted between 1984 and 1996 have demonstrated poor agreement not only between referring physicians and neurologists,\(^5\) but also among neurologists interviewing the same patients.\(^6\,^8\) Despite these limitations, the judgment of a vascular neurologist or general neurologist remains by default the gold standard for TIA diagnosis in most studies.

Since that time, the development of TIA clinics run by stroke-trained physicians has been a major improvement in the management of TIA. Evaluation and management in a TIA clinic has been associated with a significant reduction in the risk of subsequent stroke compared with prior observational studies.\(^3\)

With this as a background, we investigated among fellowship-trained stroke neurologists the level of agreement for the adjudication of the likelihood of a cerebrovascular etiology of transient neurological events based on a review of clinic notes from the Stanford TIA clinic. We also estimated the impact of different scales used to assess TIA likelihood on raters’ judgments. We hypothesized that the number of items that composes a scale might influence the judgment of individual rater on TIA likelihood and also the interrater agreement.

Materials and Methods
Between July 2006 and June 2007, 55 patients were referred from the Stanford emergency department to the Stanford TIA clinic for...
rapid outpatient evaluation. These patients were all seen in the Stanford emergency room by a neurology resident and later in the TIA clinic by a fellowship-trained stroke neurologist (none of whom participated in this study). Patient data were collected retrospectively under a Stanford Institutional Review Board-approved waiver of consent.

Notes from the initial TIA clinic visit were identified for all patients. All the patients were evaluated (and previously documented in the records) within 72 hours of their initial symptom onset. The clinical notes were reviewed by each reviewer after an average delay of 1 year after the event. Table 1 summarizes the information systematically collected from the clinical notes. All identifying information, laboratory, imaging data, assessments, and outcomes were removed. The notes were then presented to 3 fellowship-trained stroke neurologists (reviewers) who had never met these patients.

Two of the reviewers were trained in certified vascular neurology programs in the United States and the third in a similar program in France. Each reviewer was asked to rate the likelihood that the patient had a transient cerebrovascular event. Each patient was rated using 3 scales: first, a 2-point scale with choices of “likely” or “unlikely”; second, a 3-point scale with choices of “likely,” “possible,” or “unlikely”; and third, a 4-point scale with choices of “very likely,” “likely,” “possible,” and “unlikely.” Each reviewer was blinded to the scores of other reviewers.

For each scale, we assessed the overall agreement and the agreement coefficient (AC1). The AC1 is not affected by the scale type on the rate of unlikely TIA among each rater using the Holm-Bonferroni method and the level at α<0.05 was considered as significant. We then estimated the effect of the scale type on the rate of unlikely TIA among each rater using the McNemar test.

Statistical analysis was done using SAS 9.1.3 (SAS Institute Inc, Cary, NC), SPSS 17.0 (SPSS, Chicago, Ill), INTER_RATER.MAC SAS macro by K. Gwet, and MH program Version 1.2 by J. Uebersax (http://john-uebersax.com/stat/mh.htm).

Results

Agreement

The diagnoses made by each reviewer using each scale are presented in Table 2.

For the 2-point scale, the overall agreement was 0.72, AC1 0.46 (0.30 to 0.63), and ICC 0.44 (0.27 to 0.60).

For the 3-point scale, the overall agreement was 0.56, AC1 0.35 (0.21 to 0.49), and ICC 0.56 (0.41 to 0.70). When we combined the likely and possible scores and compared this with unlikely, the overall agreement was 0.66, AC1 0.34 (0.16 to 0.51), and ICC 0.58 (0.35 to 0.74). Compared with the original scale, the overall agreement improved (P=0.008), but agreement coefficients did not.

For the 4-point scale, overall agreement was 0.47, AC1 0.31 (0.20 to 0.41), and ICC 0.56 (0.38 to 0.70). We combined very likely, likely, and possible scores and compared with unlikely, the overall agreement was 0.78, AC1 0.69 (0.57 to 0.81), and ICC 0.57 (0.33 to 0.73). This combination resulted in an improvement of the overall agreement and AC1 of this scale (P<0.001). This improvement occurred because of an increase in the agreement in the positive TIA diagnoses (AC1=0.81 [0.62 to 0.99]), whereas AC1 for “unlikely TIA” remained very low (0.16 [0 to 0.65]) and not statistically significant (P=0.25).

Rating Bias

Raters’ Bias

Difference in rating bias among all raters was significant for each scale: P=0.001, 0.012, and <0.001. Paired raters analysis revealed that Reviewer 2 more often used unlikely on the 2-point scale and the 3-point scale, and Reviewer 3 categorized cases as unlikely less frequently and likely or very likely more frequently on the 4-point scale.

Table 2. Ratings and Agreement on TIA Diagnosis Using 2-, 3-, and 4-Point Scale

<table>
<thead>
<tr>
<th>Reviewer</th>
<th>Likely/Unlikely Percent (n)</th>
<th>Likely/Possible/Unlikely Percent (n)</th>
<th>Very Likely/Likely/Possible/Unlikely Percent (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>44% (24)/56% (31)</td>
<td>29% (16)/36% (20)/35% (19)</td>
<td>9% (5)/26% (14)/38% (21)/27% (15)</td>
</tr>
<tr>
<td>2</td>
<td>25% (14)/75% (41)</td>
<td>20% (11)/25% (14)/55% (30)</td>
<td>7% (4)/18% (10)/55% (30)/20% (11)</td>
</tr>
<tr>
<td>3</td>
<td>51% (28)/49% (27)</td>
<td>20% (11)/45% (25)/35% (19)</td>
<td>17% (9)/36% (20)/38% (21)/9% (5)</td>
</tr>
<tr>
<td>Raw agreement</td>
<td>0.72</td>
<td>0.56</td>
<td>0.46</td>
</tr>
<tr>
<td>AC1 value</td>
<td>0.46 (95% CI, 0.30–0.63)</td>
<td>0.35 (95% CI, 0.21–0.49)</td>
<td>0.31 (95% CI, 0.20–0.41)</td>
</tr>
<tr>
<td>ICC</td>
<td>0.44 (95% CI, 0.27–0.60)</td>
<td>0.56 (95% CI, 0.41–0.70)</td>
<td>0.56 (95% CI, 0.38–0.70)</td>
</tr>
</tbody>
</table>
Scale Bias
For each reviewer, the rate of “unlikely TIA” progressively decreased when the number of items that compose the scale increased: Reviewer 1: 2 points = 56%, 3 points = 35% (P = 0.035), and 4 points = 27% (P = nonsignificant); Reviewer 2: 2 points = 75%, 3 points = 55% (P = 0.046), and 4 points = 20% (P < 0.001); and Reviewer 3: 2 points = 49%, 3 points = 34% (P = nonsignificant), 4 points = 9% (P = 0.003). Fifty-three percent of the possible cases on the 3-point scale and 93% on the 4-point scale came from the unlikely group on the 2-point scale.

In addition, agreement on the unlikely score estimated by AC1 also decreased when there were a larger number of options on the scale: 0.55 (0.19 to 0.92) for the 2-point scale, 0.39 (0.03 to 0.75) for the 3-point scale, and 0.25 (0.19 to 0.68) for the 4-point scale (P = 0.566).

In contrast, the agreement on likely scores increased between the 2-point scale and the 3-point scale and then decreased again. Likely score on the 2-point scale was 0.33 (–0.19 to 0.85), 3-point scale = 0.57 (–0.03 to 1), and 4-point 0.20 (–0.08 to 0.49) (P = 0.547). Change in the agreement on the “possible TIA” diagnosis from the 3- to the 4-point scale was 0.17 (–0.11 to 0.45) to 0.39 (0.14 to 0.63). Agreement on the “very likely” diagnosis was 0.35 (–0.28 to 0.98) (P = 0.421).

Discussion
Our results demonstrate that TIA diagnosis based only on evaluation of a clinical note generated by another stroke neurologist results in moderate agreement among fellowship-trained stroke neurologists. These results also show the subjectivity of the diagnostic impressions based on different raters’ “thresholds” for making a diagnosis of likely versus unlikely TIA. We further showed that the type of scale used can influence the ratings and that the degree of disagreement on the “unlikely TIA” diagnosis is substantial and may have important implications for patient management.

Previous studies have shown a low rate of agreement between nonneurologists and neurologists as well as between neurologists. Based on clinical reports, without the help of direct interview and access to vessel and brain imaging, our study found only a moderate agreement between physicians trained at different institutions. All scales had a low level of agreement ranging from fair to moderate and the level of agreement varied for the specific diagnosis of “unlikely TIA” when the number of additional options on the scale was increased. The combination of positive TIA findings (likely, possible, very likely) improved the performance of this scale in the agreement on positive diagnoses. However, agreement remained poor for the diagnosis “unlikely TIA.” These findings have important implications: First, in clinical practice, in which the urgency of the initial evaluation depends on physician perception of TIA likelihood, patients thought unlikely to have had a TIA may be less likely to have an expedited evaluation. Second, these patients would be excluded from many prospective studies that report stroke risk only among patients diagnosed with a “likely” or “possible” TIA. For example, in our study, up to 70% of cases were in the “unlikely” category based on the 2-point scale performed by 1 rater, whereas only 9% would have been considered “unlikely” by another rater using the 4-point scale. These findings argue for the use of more objective measures of TIA likelihood for both stroke specialists as well as primary care and emergency medicine specialists. Recent studies have suggested that ABCD² score items and other scales could be a useful tool to confirm the diagnosis of TIA. Several studies have suggested that acute MRI can also be useful to demonstrate the vascular nature of transient neurological symptoms. The diffusion-weighted imaging sequence is positive in approximately 39% of patients with TIA with a frequency by site ranging from 25% to 67%. In addition, perfusion imaging can increase the yield of MRI because up to 50% of patients with TIA had either a positive diffusion-weighted image or perfusion-weighted imaging in 1 recent study. These findings suggest that both a structured report and multimodal MRI results are tools that can improve diagnostic reliability.

Study Limitations
Our study has limitations. First, the number of cases was quite low, and our findings remain to be validated on a larger data set. Second, we did not report the conclusion of the neurologist who treated the patient. We made this choice because our investigations focused on the agreement among reviewers who had no knowledge of the imaging results or patient outcomes.

In conclusion, our results suggest that the interpretation of TIA likelihood remains highly subjective. A structured report, including a systematic review of pertinent symptoms and relevant medical history as well as acute brain imaging with MRI, may improve the reliability of the diagnosis of TIA. Future studies are warranted to evaluate the yield of these tools for the diagnosis of transient neurological symptoms in consecutive patients.

Disclosure
None.

References


Agreement Regarding Diagnosis of Transient Ischemic Attack Fairly Low Among Stroke-Trained Neurologists
James Castle, Michael Mlynash, Karmung Lee, Anna Finley Caulfield, Connie Wolford, Stephanie Kemp, Scott Hamilton, Gregory W. Albers and Jean-Marc Olivot

Stroke. 2010;41:1367-1370; originally published online May 27, 2010; doi: 10.1161/STROKEAHA.109.577650

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/41/7/1367

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2016/03/31/STROKEAHA.109.577650.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org/subscriptions/
背景和目的：内科医师对短暂性脑缺血发作（transient ischemic attack, TIA）诊断的一致性差。已有研究对神经科医师与非神经科医师，神经科医师之间诊断的一致性做了对比，但还未在经过中风专业培训的神经科医师中间进行这种比较。故本研究就 55 例疑似 TIA 的患者对这些专业医师的诊断一致性进行了调查研究。

方法：由斯坦福急诊转入斯坦福门诊的 55 例患者，分别由 3 位接受过卒中专业培训的神经科医师查看相关病史和体格检查记录，并予以盲法评价。这 3 位医师事前不了解患者的任何检查结果及预后情况。根据临床表现与 TIA 的相似程度，将每个患者诊断为 TIA 的可能性进行分级。我们采用了 3 种不同的分级系统（分别是 2 分法，3 分法，4 分法）。我们评价了评价者间的整体一致性，同时评价了与评价个体和分级系统相关的偏倚。

结果：尽管我们用统计学方法对所采用的分级系统进行评价，排除所采用的分级系统及统计检验方法的影响，这些接受过卒中专业培训的神经科医师对 TIA 诊断的一致性仍较差。对每种分级方法而言，所有评价者中评价偏倚的差异是明显的，P 值分别为 0.001，0.012 和 <0.001。此外，对于每个评价者来说，随着分级级别增加，标记为“不可能是 TIA”的诊断率在逐渐下降。

结论：即使在卒中专业医师中间，TIA 的诊断仍然具有很强的主观性，在临床评价之外，仍然需要各种检查来验证诊断的正确性，即使在诊断可能 TIA 时也应非常谨慎。

关键词：诊断，短暂性脑缺血发作，评价者间一致性

| 表 1 提供给评价者的患者信息
| 主诉和现病史* |
| 常规检查和用药情况* |
| 社会史* |
| 家族史* |
| 系统回顾* |
| 最初生命体征* |
| 体格检查（包括完整的神经系统检查）* |
| * 是由发病时负责接诊的内科医师记录在首次评价中的内容，去除了所有诊断性影像学结果和最后的初步诊断。

| 表 2 采用 2 分法，3 分法，4 分法分级系统进行分级和对 TIA 诊断的一致性
<table>
<thead>
<tr>
<th>评价者</th>
<th>可能 TIA/ 不可能 TIA</th>
<th>可能 / 疑似 / 不可能 TIA</th>
<th>极可能 / 可能 / 疑似 / 不可能 TIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>44% (24)/56% (31)</td>
<td>29% (16)/36% (20)/35% (19)</td>
<td>9% (5)/26% (14)/38% (21)/27% (15)</td>
</tr>
<tr>
<td>2</td>
<td>25% (14)/75% (41)</td>
<td>20% (11)/25% (14)/55% (30)</td>
<td>7% (4)/18% (10)/55% (30)/20% (11)</td>
</tr>
<tr>
<td>3</td>
<td>51% (28)/49% (27)</td>
<td>20% (11)/45% (25)/35% (19)</td>
<td>17% (9)/36% (20)/38% (21)/9% (5)</td>
</tr>
<tr>
<td>相关一致性</td>
<td>0.72</td>
<td>0.56</td>
<td>0.46</td>
</tr>
<tr>
<td>AC1 值</td>
<td>0.46 (95% CI, 0.30-0.63)</td>
<td>0.35 (95% CI, 0.21-0.49)</td>
<td>0.31 (95% CI, 0.20-0.41)</td>
</tr>
<tr>
<td>ICC</td>
<td>0.44 (95% CI, 0.27-0.60)</td>
<td>0.56 (95% CI, 0.41-0.70)</td>
<td>0.56 (95% CI, 0.38-0.70)</td>
</tr>
</tbody>
</table>

AC1，一致系数；ICC，组内相关。