Substantial Observer Variability in the Differentiation Between Primary Intracerebral Hemorrhage and Hemorrhagic Transformation of Infarction on CT Brain Imaging

Caroline E. Lovelock, FRACP; Philip Anslow, FRCR; Andrew J. Molyneux, FRCR; James V. Byrne, FRCR; Wilhelm Kuker, FRCR; Pieter M. Pretorius, FRCR; Andrew Coull, MD; Peter M. Rothwell, FMedSci

Background and Purpose—CT remains the most commonly used imaging technique in acute stroke but is often delayed after minor stroke. Interobserver reliability in distinguishing hemorrhagic transformation of infarction from intracerebral hemorrhage may depend on delays to CT but has not been reported previously despite the clinical importance of this distinction.

Methods—Initial CT scans with intraparenchymal hematoma from the first 1000 patients with stroke in the Oxford Vascular Study were independently categorized as intracerebral hemorrhage or hemorrhagic transformation of infarction by 5 neuroradiologists, both blinded and unblinded to clinical history. Thirty scans were reviewed twice. Agreement was quantified by the $\kappa$ statistic.

Results—Seventy-eight scans showed intraparenchymal hematoma. Blinded pairwise interrater agreements for a diagnosis of intracerebral hemorrhage ranged from $\kappa=0.15$ to $0.48$ with poor overall agreement ($\kappa=0.35$; 95% CI, 0.15 to 0.54) even after unblinding ($\kappa=0.41$; 0.21 to 0.60). Blinded intrarater agreements ranged from $\kappa=0.21$ to 0.92. Lack of consensus after unblinding was greatest in patients scanned $\geq 24$ hours after stroke onset (67% versus 25%, $P=0.001$) and in minor stroke (National Institutes of Health Stroke Scale $\leq 5$: 56% versus 29%, $P=0.04$) with disagreement in 75% of patients scanned $\geq 24$ hours after minor stroke and in 48% of all 30-day stroke survivors in whom reliable diagnosis would be expected to influence long-term management.

Conclusion—Reliability of diagnosis of intraparenchymal hematoma on CT brain scan in minor stroke is poor, particularly if scanning is delayed. Immediate brain imaging is justified in patients with minor stroke. (Stroke. 2009;40:3763-3767.)

Key Words: computerized tomography • interobserver reliability • intracerebral hemorrhage

Primary intracerebral hemorrhage (ICH) accounts for 10% to 15% of all strokes and affects around 23 000 patients in the United Kingdom and 60 000 in the United States per year. ICH must be distinguished from ischemic stroke because the pathogenesis, treatment, and prognosis are substantially different. Although MRI is superior at detecting ICH that is $>1$ week old, CT has been the gold standard for identifying acute ICH and remains the most commonly used brain imaging modality for acute stroke. However, the diagnosis of acute ICH on CT is not necessarily straightforward, because the appearance of hemorrhagic transformation of infarction (HTI) can easily resemble ICH, and there are no agreed radiological criteria for differentiating between these pathological subtypes. Although interobserver reliability in the CT differentiation between HTI and infarction has been assessed, there have been no published reports of the more clinically important differentiation between HTI and ICH. Spontaneous HTI occurs within the first few days in up to 15% of ischemic strokes. The appearance of HTI can be broadly classified as either infarction with petechial hemorrhage or intrainfarct hematoma, and this latter type can be confused with ICH radiologically. HTI probably occurs when the hypoxic vascular bed in the infarcted region is exposed to the force of arterial blood pressure through recanalized vessels. Given that HTI begins with arterial occlusion, most clinicians would start anticoagulation or antiplatelet therapy as secondary stroke prevention after the hemorrhage has resolved. However, most clinicians would be reluctant to use these therapies after ICH unless the risk of thromboembolism was also high given that the risk of

Received April 27, 2009; final revision received June 29, 2009; accepted July 10, 2009.
From the Stroke Prevention Research Unit (C.E.L., A.C., P.M.R.), Department of Clinical Neurology, John Radcliffe Hospital, Oxford, UK; and the Department of Neuroradiology (P.A., A.M., J.V.B., W.K., P.M.P.), John Radcliffe Hospital, Oxford, UK.
Correspondence to Peter M. Rothwell, FMedSci, Stroke Prevention Research Unit, University Department of Neurology, Level 6, West Wing, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK. E-mail peter.rothwell@cneuro.ox.ac.uk
© 2009 American Heart Association, Inc.

Stroke is available at http://stroke.ahajournals.org

DOI: 10.1161/STROKEAHA.109.553933
recurrent hemorrhage outweighs the risk of a future ischemic stroke.7

Because the incidence of HTI is low within the first 24 hours after stroke onset but increases thereafter,8–10 and because ICH may come to resemble HTI as hematoma density decreases with time,11 delays in brain imaging are likely to make the distinction between ICH and HTI more difficult. Yet, despite efforts to minimize scan delays for patients who are thrombolysis candidates, other patients with minor stroke symptoms, elderly patients, or those who have presented outside the therapeutic time window for thrombolysis are still likely to face scan delays. In the United Kingdom, up to half of patients with stroke presenting outside normal working hours are scanned >24 hours after stroke onset,12 and in some US centers, significant numbers of patients present to emergency departments >12 to 24 hours poststroke onset13,14 or are not scanned immediately on arrival to the hospital.15,16 We therefore determined interobserver reliability in the diagnosis of ICH versus HTI stratified according to delay from onset of stroke to imaging with particular emphasis on interpretation of brain imaging for cases who subsequently survived >30 days poststroke for whom a lack of reliability in radiological diagnosis would have the most impact on decisions over the most appropriate secondary prevention therapy.

Methods

All initial non-contrast CT scans showing any intraparenchymal hematoma from the first 1000 consecutive cases of stroke in the Oxford Vascular Study (OXVASC) were reviewed. OXVASC is a population-based study of all acute vascular events, including stroke, in a population of 91,000 individuals registered with 63 general practitioners. The methods of case ascertainment and clinical assessment in OXVASC have been described in detail previously.17 Scans showing minor petechial hemorrhage in association with infarction were not included because no controversy regarding stroke etiology from the scan appearance was expected. Scans of subarachnoid hemorrhage, traumatic intracranial bleeds, and hemorrhages into tumors were also excluded.

Five consultant neuroradiologists independently reviewed the films and diagnosed the hematoma as either ICH or HTI. Because a diagnostic decision is almost always required in routine clinical practice for planning further investigations and choosing secondary prevention therapy, no “uncertain” category was permitted. The observers were initially blinded to the clinical history and then given the opportunity to change the radiological diagnosis in light of relevant clinical information, including exposure to anticoagulation therapy and delay from ictus to scan. All radiologists reviewed 30 films twice after an interval of 12 months blinded to their original assessment so that intrarater variability could be assessed.

Other imaging data were also collected, including the volume and location of the hematoma. Hemorrhage volumes were measured using the ABC/2 method.18 The location of the hemorrhage was categorized as lobar or deep/posterior if it involved the basal ganglia, thalamus, or cerebellum and brainstem. Clinical and demographic data were collected by interviewing patients and/or relatives shortly after clinical presentation and also by reviewing hospital and general practitioner records. Stroke severity was assessed using the National Institutes of Health Stroke Scale (NIHSS), and surviving patients were followed up at 1 month and assessed for their level of disability using the modified Rankin Scale score. We reviewed the records of cases in which there was disagreement over the radiological diagnosis to ascertain the diagnosis made by the treating physicians at the time the initial CT brain scan was performed and the subsequent choice of secondary prevention therapy.

Statistical Analysis

Because the rate of HTI is relatively low within the first 24 hours after stroke onset,9,10 hemotoma density in ICH decreases with time,11 and because we expected greater radiological agreement in those cases with severe strokes due to large-volume intracerebral hematomas who are more likely to present early and be scanned immediately, we stratified the analysis according to whether or not the scan was performed <24 of stroke onset. Observer agreement was measured using unweighted κ statistics. We also determined the proportion of cases in which there was complete agreement among all 5 neuroradiologists (consensus).

Comparisons between groups with and without a consensus diagnosis of ICH were made using Fisher exact test, the t test, and the Mann–Whitney U test. All analyses were performed using SPSS Version 12 statistical software package and STATA.

Results

Of the first 1000 consecutive cases of stroke ascertained in OXVASC from April 2002 to October 2006, we identified 98 patients with any intraparenchymal hemorrhage on the initial CT brain scan. Of these patients, we excluded 9 with petechial HTI, 2 with tumor-related bleeds, 3 with traumatic ICH, and 6 with subarachnoid hemorrhage. The 78 remaining patients with intraparenchymal hematomas (56% male, mean age 76 years) were included in our analysis. The exact onset of stroke could not be determined in 2 patients whose stroke occurred during prolonged anesthesia in the intensive care unit after admission for a nonstroke illness. In the remaining 76 patients, the median delay (interquartile range) from ictus to CT scan was 7 hours (3 to 40 hours) and from presentation to secondary care and scan was 2.5 hours (1 to 10.5 hours).

There were 26 patients who were scanned at ≥24 hours after stroke onset, of whom 4 had presented 12 to 24 hours after stroke onset and 9 had presented ≥24 hours.

Pairwise interrater agreements for a diagnosis of HTI versus ICH on the 78 initial scans with nontraumatic intraparenchymal hematoma were poor to moderate, ranging from κ=0.15 to 0.48. The overall interrater agreement between blinded observers was poor (κ=0.35; 95% CI, 0.15 to 0.54) and did not substantially improve with unblinding (κ=0.41; 0.21 to 0.60). With the benefit of all relevant clinical and radiological information, all 5 observers diagnosed HTI in 4 cases (5%) and ICH in 46 cases (59%), but there was lack of consensus across the 5 observers in 28 (36%) cases. Examples of scans in which there was a lack of consensus are shown in the Figure.

When 30 scans from the first half of the study were reviewed again, intrarater reproducibility was variable for blinded (κ1=0.21, κ2=0.51, κ3=0.52, κ4=0.57, κ5=0.92) and unblinded observations (κ1=0.36, κ2=0.37, κ3=0.52, κ4=0.65, κ5=0.92).

The Table compares clinical and radiological data for the groups with and without consensus agreement for a diagnosis of ICH among all 5 neuroradiologists after unblinding to the clinical history (ie, excluding the 4 cases that were unanimously diagnosed as HTI). As expected, patients without a consensus diagnosis of ICH tended to have less severe strokes as shown by the lower median NIHSS score at presentation (6 versus 17, P=0.04) and by the higher proportion with good outcomes (modified Rankin Scale score <3) at 30 days (46% versus 22%, P=0.04). In the group without a consensus
diagnosis of ICH, the median hematoma volume (mL) was smaller compared to the group with a consensus diagnosis (7 versus 27, \(P = 0.04\)), the median delay (hours) from ictus to scan was longer (33 versus 4, \(P = 0.001\)), and fewer cases were scanned within 24 hours (43% versus 83%, \(P = 0.001\)). Other factors associated with lack of a consensus diagnosis included older age (\(P = 0.01\)) and lobar location of bleed (\(P = 0.02\)). There was lack of consensus in 55% (17 of 31) of patients with a lobar bleed, in 47% (17 of 36) of patients aged \(\geq 80\) years, and in 69% (9 of 13) aged \(\geq 80\) years with a lobar bleed.

When stratifying radiological agreement by delay to scan, lack of consensus was more likely in scans performed \(\geq 24\) hours after stroke onset whether blinded to clinical history (58% versus 25%, \(P = 0.01\)) or unblinded (67% versus 25%, \(P = 0.001\)). This difference remained even when the unblinded analysis was limited to patients with minor stroke (NIHSS \(\leq 5\)) in whom there was a lack of radiological agreement in 12 of 16 cases scanned at \(\geq 24\) hours versus 2 of 9 scanned within 24 hours (\(P = 0.017\)). Patients scanned \(\geq 24\) hours after stroke onset had lower median NIHSS scores (4 versus 19, \(P < 0.001\)) and were more likely to survive beyond 30 days (75% versus 38%, \(P = 0.002\)) and therefore require long-term secondary prevention.

Of those cases alive at 30 days, for whom a change in diagnosis would usually result in a change in secondary prevention, there was a lack of consensus in 19 of 40 (48%) cases. At the time of presentation, the treating team diagnosed 11 of these cases as ICH and discontinued antithrombotic therapy, although 5 cases had atrial fibrillation or known atheroembolic disease. In 2 cases, the treating team diagnosed HTI and initiated aspirin; one of these cases had a further ICH within 6 months. In 6 cases, all of whom had risk factors for atheroembolic disease, the treatment team was unsure of the diagnosis. Antithrombotic therapy was stopped in 2 of these cases, one of whom had a subsequent ischemic stroke within 6 months.

**Discussion**

We have shown that interrater reproducibility in differentiating between ICH and HTI in patients with acute stroke with hematoma on initial CT is only moderate at best and that intrarater reproducibility is variable. Lack of agreement partly reflects a lack of generally agreed radiological criteria for distinguishing between these 2 stroke subtypes. Importantly, the lack of interobserver agreement is greatest among patients scanned at \(\geq 24\) hours, particularly when presenting with minor stroke syndromes. Patients with minor stroke may be scanned late because they present late or because immediate imaging is not considered necessary if they are not candidates for thrombolysis. However, earlier scanning may improve the reliability of radiological diagnosis in this group who are more likely to survive and in whom the distinction between HTI and ICH is of most relevance, because it has implications for prognosis and choice of secondary stroke prevention.

Our study design, in which all cases with hematoma on an initial CT scan were prospectively identified from a population-based study, avoided selection bias. A retrospective study in which scans were chosen on the basis of the initial radiological diagnosis might have missed cases that were classified as ischemic stroke rather than HTI, and an emergency department-based study might have missed cases with minor stroke due to an ICH and a delayed presentation, which form the group in which there was least radiological consensus.

Awareness of the potential of HTI to mimic ICH was raised 18 years ago in a study of 15 patients with stroke with no hemorrhage on an initial scan performed within 6 hours and hemorrhage on a second scan performed 18 hours later. All second scans were thought to show ICH by both neurologists and neuroradiologists blinded to the earlier scan results.\(^1\) Although awareness of HTI is now much higher, ours is the first published report of interobserver agreement in the differentiation of HTI and ICH. Previous reports of interpretation of CT brain imaging in acute stroke have focused on observer reproducibility at the other end of the spectrum—the differentiation between HTI and infarction.\(^2,3,10\)

We identified a number of variables that were associated with a lack of radiological consensus, many of which were probably interrelated: delay to scan, older age, lobar location of hemorrhage, lower NIHSS score, and smaller hematoma volume. Delay from stroke onset to scan of \(\geq 24\) hours was

![Figure](image_url)
practice, in which decisions must be made as to the etiology were low. However, the study design reflects real life in which some delay to scan due to patient-related factors is inevitable. Moreover, delays were less than in most parts of the United Kingdom, where up to half of inpatients with stroke presenting outside normal working hours are scanned beyond 24 hours with even longer delays to outpatient CT scanning. Delays to scanning are also reported in other developed countries, including the United States, particularly in rural settings where there are often significant transport delays to accessing stroke care, and in developing nations, where there is limited access to CT imaging.

A significant level of disagreement in differentiating HTI and ICH among stroke survivors has important clinical implications. Most clinicians would start antithrombotic therapy after HTI, but not usually after ICH. In this study, antithrombotic agents were stopped or not initiated in some of the stroke survivors without a consensus diagnosis, even when there was an ongoing high risk of future ischemic stroke, including atrial fibrillation. Assuming that these cases had HTI secondary to cardioembolism, the annual risk of a further stroke without prophylaxis is approximately 12%, and there would have been a case for anticoagulating them at a later date or starting antplatelet therapy. However, if these cases had a primary ICH, then evidence from a systematic review of studies reporting recurrent strokes in survivors of ICH suggests that their annual risk of a further ICH is 2.3%, and if a lobar ICH, the annual risk is higher again at 4.4% to 13.6%. These risks are likely to be increased by taking antithrombotic drugs. In particular, warfarin use post-ICH is estimated to reduce quality-adjusted life expectancy by up to 2 years.

Our findings also have implications for clinical trials reporting ICH as an outcome. In 2 reviews of oral anticoagulants for primary and secondary prevention of stroke in patients with nonvalvular atrial fibrillation, only 2 of the 9 included trials differentiated HTI from ICH as outcomes, yet the distinction between these stroke subtypes is clinically relevant in this setting. An embolic ischemic stroke associated with hemorrhagic transformation could be thought of as an outcome that has occurred despite the antithrombotic agent and ICH as an outcome that might have occurred because of the antithrombotic agent.

There are 2 potential strategies for improving reliability of diagnosis of ICH versus HTI. The first strategy is to maximize the proportion of patients undergoing brain imaging within 24 hours of stroke onset, because rates of consensus are higher when the delays to CT scanning are <24 hours, and this is particularly relevant for patients presenting with minor stroke. A hematoma identified within the first 24 hours of stroke onset is more likely to be ICH, because very early HTI with hematoma formation is relatively uncommon.

### Table. Comparison of Clinical and Radiological Findings Between Groups With and Without a Complete Consensus Diagnosis of ICH After Unblinding to Clinical History

<table>
<thead>
<tr>
<th>Clinical factors</th>
<th>No Consensus Diagnosis (N=28)</th>
<th>Consensus Diagnosis of ICH (N=46)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (SD)</td>
<td>80 (10)</td>
<td>73 (13)</td>
<td>0.01</td>
</tr>
<tr>
<td>Males</td>
<td>13 (46%)</td>
<td>29 (63%)</td>
<td>0.23</td>
</tr>
<tr>
<td>Median NIHSS (IQR)</td>
<td>6 (3–17)</td>
<td>17 (6–30)</td>
<td>0.04</td>
</tr>
<tr>
<td>NIHSS ≥5</td>
<td>14 (50%)</td>
<td>11 (24%)</td>
<td>0.04</td>
</tr>
<tr>
<td>History of hypertension</td>
<td>11 (39%)</td>
<td>20 (43%)</td>
<td>0.81</td>
</tr>
<tr>
<td>History of atrial fibrillation</td>
<td>7 (25%)</td>
<td>6 (13%)</td>
<td>0.22</td>
</tr>
<tr>
<td>Previous stroke</td>
<td>7 (25%)</td>
<td>5 (11%)</td>
<td>0.19</td>
</tr>
<tr>
<td>Premorbid warfarin use</td>
<td>3 (11%)</td>
<td>8 (17%)</td>
<td>0.52</td>
</tr>
<tr>
<td>Premorbid aspirin use</td>
<td>11 (39%)</td>
<td>16 (35%)</td>
<td>0.81</td>
</tr>
<tr>
<td>Radiological factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median volume (IQR) of hematoma, mL</td>
<td>7 (3–34)</td>
<td>27 (7–62)</td>
<td>0.04</td>
</tr>
<tr>
<td>Volume ≤10 mL</td>
<td>15 (54%)</td>
<td>12 (26%)</td>
<td>0.03</td>
</tr>
<tr>
<td>Volume ≤20 mL</td>
<td>19 (68%)</td>
<td>20 (43%)</td>
<td>0.06</td>
</tr>
<tr>
<td>Median delay (IQR) from ictus to scan, hours</td>
<td>33 (5–74)</td>
<td>4 (3–16)</td>
<td>0.001</td>
</tr>
<tr>
<td>Delay to scan ≥12 hours</td>
<td>18 (64%)</td>
<td>12 (27%)</td>
<td>0.003</td>
</tr>
<tr>
<td>≥24 hours</td>
<td>16 (57%)</td>
<td>8 (18%)</td>
<td>0.001</td>
</tr>
<tr>
<td>≥48 hours</td>
<td>11 (40%)</td>
<td>2 (5%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Location of hematoma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lobar</td>
<td>17 (61%)</td>
<td>14 (30%)</td>
<td>0.015</td>
</tr>
<tr>
<td>Deep/posterior fossa</td>
<td>11 (39%)</td>
<td>32 (70%)</td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alive at 30 days</td>
<td>19 (68%)</td>
<td>20 (43%)</td>
<td>0.06</td>
</tr>
<tr>
<td>Modified Rankin Scale score ≥3 at 30 days</td>
<td>13 (46%)</td>
<td>10 (22%)</td>
<td>0.04</td>
</tr>
</tbody>
</table>

*Excludes 4 cases with complete consensus for diagnosis of HTI. IQR indicates interquartile range.
The second strategy is to consider MRI brain imaging for patients who present later than this, particularly for those with lobar hematomas because the recurrence rate of primary lobar ICH is relatively high, and for those who require treatment with anticoagulation. The interobserver reliability for differentiating between HTI and ICH using MRI has not yet been shown; a number of MRI sequences may provide additional useful information that make the distinction easier compared with CT. Gradient echo MRI can accurately identify intracerebral hemorrhage even in the hyperacute stage. Diffusion-weighted images combined with apparent diffusion coefficient maps can identify a wider region of infarction surrounding the hematoma in HTI; and although apparent diffusion coefficients are sometimes reduced within the rim of restricted diffusion that surrounds primary hematomas suggestive of ischemic injury, these values rapidly rise which does not occur in regions of infarction. Furthermore, diffusion-weighted imaging can identify contemporaneous areas of infarction in other vascular territories suggestive of cardiac embolism.

In conclusion, differentiation between HTI and ICH on CT scan in acute stroke has only modest reproducibility, even with knowledge of all relevant clinical information. This finding has important implications for clinical practice, epidemiological studies, and CT-based clinical trials. CT appearances of intraparenchymal hematomas are more reliably interpreted if scanning is performed within 24 hours, but less so beyond this time. This finding has particular relevance to the management of patients with minor stroke who often present late and are therefore scanned late.

Sources of Funding
The Oxford Vascular Study is funded by the UK Medical Research Council, the National Institute of Health Research, the Stroke Association, the Dunhill Medical Trust, and the Oxford Partnership Comprehensive Biomedical Research Centre.

Disclosures
None.

References