Observer variation in the interpretation of intra-arterial angiograms and the risk of inappropriate decisions about carotid endarterectomy

G R Young, P A G Sandercock, J Slattery, P R D Humphrey, E T S Smith, L Brock

Abstract

Objective—To determine how often observer variation in the interpretation of intra-arterial angiograms might alter the decision whether or not to refer patients for carotid surgery.

Methods—A prospective study was carried out in a consecutive series of 99 patients with transient ischaemic attacks and minor strokes. Interpretable angiographic films were available for 179 carotid artery bifurcations. Stenosis of the internal carotid artery was measured using mm scales, independently by three different radiologists (A, B, and C), using the European Carotid Surgery Trial method.

Results—An analysis of the grouped data showed good to moderate agreement by kappa statistics for radiologists A v B, B v C, and A v C of 0.68, 0.60, and 0.70 respectively. The absolute difference in the estimate of stenosis by each of the different radiologists (interobserver variation) was 9.5% and for each radiologist on two separate occasions (intraobserver variation) 8.4%. The degree of observer error was smallest among severely stenosed arteries. Although the absolute differences were small, “clinically important” differences which could change the treatment recommended from surgery to no surgery (or vice versa) occurred between radiologists A and B, B and C, and A and C in: seven (3.9%) six (3.4%), and 11 (6.1%) vessels respectively.

Conclusions—Because observer variation affects all of the imaging methods (Doppler, duplex, contrast arteriography, and MR angiography) used to select patients with transient ischaemic attack and stroke, these findings are likely to be widely relevant. Any centre assessing patients with cerebrovascular disease will need to implement strict quality control measures in the interpretation of angiograms (and other vascular imaging procedures) to minimise observer error and thereby reduce the number of inappropriate decisions made to refer for carotid artery surgery or not.

(J Neurol Neurosurg Psychiatry 1996;60:152–157)

Keywords: angiography; carotid stenosis; diagnosis

The publication of the results of the European Carotid Surgery Trial (ECST) and the North American Symptomatic Carotid Endarterectomy Trial (NASCET) has clarified the place of surgery in the management of patients with transient ischaemic attacks and minor stroke who have stenosis of the ipsilateral internal carotid artery bifurcation.1 2 In general, patients with severe stenosis (70–99%) are likely to derive substantial benefit from surgery, whereas for patients with mild stenosis (0–29%), the balance of risks and benefits will, in general, be against surgery. Uncertainty persists for patients with intermediate degrees of stenosis (30–69%) and awaits further publications from the ECST and NASCET studies. The implications of these results for the interpretation of cerebral angiograms are therefore substantial. In patients with transient ischaemic attacks or minor strokes, small differences in interpretation might make the difference between the same patient being referred for carotid surgery (with its attendant 4–5% risk of stroke or death) by one observer and selected for a conservative “no surgery” policy by another. The present study was undertaken in a regional neurosciences unit with the most advanced intra-arterial digital subtraction angiography and highly trained neuroradiologists, to determine to what extent intraobserver variation might lead to important differences in clinical decisions about surgery.

Materials and methods

PATIENTS AND MEASUREMENT OF STENOSIS

The data reported were collected as part of a study to examine the diagnostic accuracy of duplex carotid ultrasound.3 A consecutive series of 99 patients with carotid distribution transient ischaemic attacks or minor stroke, who were being considered for carotid surgery, underwent both duplex sonography and intra-arterial digital subtraction angiography. Anteroposterior, lateral and, where necessary, oblique views of the carotid bifurcation were obtained after selective catheterisation of the common carotid artery. Copies of the angiograms were made, and were then presented to three radiologists (A, B, and C) independently. The radiologists were not aware of the clinical details of the patients, the findings of the other radiologists, or the findings of the duplex ultrasound. They were asked to estimate the degree of stenosis of the
internal carotid artery just above the bifurcation, using the standard method outlined in
the protocol of the European Carotid Surgery Trial. This required two measurements, the
diameter of the residual lumen of the internal carotid artery at the point of maximal stenosis,
and the diameter of the original lumen of the vessel, at the same point. The contour of the
original vessel wall cannot be directly visualised by digital subtraction angiography and
so its position was estimated by the reporting radiologist. The degree of stenosis was cal-
culated by subtracting the diameter of the residual lumen from that of the original lumen and
expressing the result as a percentage of the original lumen. Measurements were made
from the films using a scale graduated in mm. Each angiogram was assessed independently
by each radiologist on two separate occasions, with an interval of about 12 months between
assessments.

STATISTICAL METHODS
The differences between the first and second readings of each radiologist were calculated for
each of the angiograms. These differences were plotted against the means of the two
readings for each angiogram to give a visual impression of the agreement and any relation
to the degree of stenosis being measured. Perfect agreement would be represented by all
points lying on a straight line through 0% disagreement; the greater the scatter of points,
the less the agreement. Any systematic bias, such as one of the radiologists consistently
reporting more severe stenosis for one of the readings, would result in a mean of the differ-
ences significantly different from zero. Similar plots were constructed between each of the
three possible pairings of radiologists to show the interobserver agreement in reporting.
When the differences between measurements are normally distributed, the 95% limits of
agreement can be calculated as the range of values encompassing the mean difference
± 2 SD.

A separate analysis was made by taking the absolute difference between pairs of measure-
ments and calculating the mean (SD). The range, median, and 95th percentile for the
absolute differences were also calculated.

The estimate of percentage stenosis was then used to classify the patients into one of four
categories for each artery: 0–29%, 30–69%, 70–99%, and completely occluded (100%). Kappa statistics for agreement were
calculated by the standard method. The kappa statistic attempts to show the propor-
tion of results assigned to the same category while taking account of any agreement that
could be expected to occur by chance. A kappa value of one indicates perfect agreement
and a value of zero indicates no agreement other than that which could be expected by chance.
Guidelines for the interpretation of kappa are often quoted as: < 0, poor agreement; 0–0.20,
slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80,
substantial agreement; 0.81–1.00 almost perfect agreement.

DEFINITION OF CLINICALLY IMPORTANT
DISAGREEMENTS
Clinically important disagreement was defined as a disagreement that was likely to have sub-
stantial clinical impact—that is, the difference between recommending surgery or not. This
analysis was performed by applying criteria derived from the interim results of the ECST
in which one radiologist classified the patient as having a degree of stenosis for which
surgery was definitely indicated (70–99%) whereas the other radiologist gave the patient a
percentage stenosis in which surgery was uncertain (30–69%) or definitely not indicated
(less than 30% or completely occluded). The proportions of angiograms in which there were
clinically significant disagreements in reporting were calculated.

Results
A total of 99 patients and 198 vessels were examined during the period of the study. Of
these, 179 vessels (90%) were reported by all three radiologists on two occasions. Angio-
grams from the remaining 19 vessels were considered by at least one radiologist on one
Table 1  Numbers of patients in each category of stenosis: radiologist A v radiologist B

<table>
<thead>
<tr>
<th>Radiologist B</th>
<th>0%-29%</th>
<th>30%-69%</th>
<th>70%-99%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiologist A:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0%-29%</td>
<td>87</td>
<td>26</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30%-69%</td>
<td>2</td>
<td>22</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>70%-99%</td>
<td>0</td>
<td>3</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>16</td>
</tr>
</tbody>
</table>

Number of clinically important disagreements leading to a difference between “surgery” and “no surgery” = 7 (3.9%), kappa = 0.68.

Table 2  Numbers of patients in each category of stenosis: radiologist A v radiologist C

<table>
<thead>
<tr>
<th>Radiologist C</th>
<th>0%-29%</th>
<th>30%-69%</th>
<th>70%-99%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiologist A:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0%-29%</td>
<td>95</td>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30%-69%</td>
<td>5</td>
<td>19</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>70%-99%</td>
<td>0</td>
<td>3</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>100%</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>14</td>
</tr>
</tbody>
</table>

Number of clinically important disagreements leading to a difference between “surgery” and “no surgery” = 8 (3.4%), kappa = 0.70.

Figures 2–3 show the agreement between the different radiologists on their first readings and the relation between the level of agreement and the degree of stenosis. The mean differences (95% confidence intervals (95% CIs)) were: $-6.4 (-8.1$ to $-4.7)$% for A–B, $-4.4 (-6.5$ to $-2.3)$% for A–C, and $2.0 (-0.1$ to $4.1)$% for B–C. Similar plots comparing the results for each radiologist reporting the same films on two separate occasions were made (figures available on request): mean differences (95% CIs) were: $0.1 (-1.2$ to $1.5)$% for A, $-2.0 (-4.2$ to $0.2)$% for B, and $1.7 (-0.7$ to $4.1)$% for C. The differences did not follow a normal distribution and therefore the 95% limits of agreement have not been calculated.

Tables 1–3 show, for each of the possible pairs of radiologists, the numbers of vessels classified to the same categories of mild, mod-
Table 3  Numbers of patients in each category of stenosis: radiologist B v radiologist C

<table>
<thead>
<tr>
<th>Radiologist C</th>
<th>0%-29%</th>
<th>30%-69%</th>
<th>70%-99%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiologist B:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0%-29%</td>
<td>79</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30%-69%</td>
<td>21</td>
<td>26</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>70%-99%</td>
<td>0</td>
<td>5</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>100%</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>13</td>
</tr>
</tbody>
</table>

Number of clinically important disagreements leading to a difference between “surgery” and “no surgery” = 11 (61%), kappa = 0.61.

Table 4  Absolute differences between the same radiologist on two readings of the same film (intraobserver variation) and between radiologists at the first reading (interobserver variation)

<table>
<thead>
<tr>
<th>Absolute differences in % stenosis</th>
<th>Range</th>
<th>Median</th>
<th>95th percentile</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differences between same radiologist on two different readings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiologist A</td>
<td>0-71</td>
<td>4</td>
<td>15</td>
<td>4.9 (7.8)</td>
</tr>
<tr>
<td>Radiologist B</td>
<td>0-63</td>
<td>7</td>
<td>33</td>
<td>9.8 (11.4)</td>
</tr>
<tr>
<td>Radiologist C</td>
<td>0-67</td>
<td>5</td>
<td>37</td>
<td>10.4 (12.7)</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td>8.4 (10.6)</td>
</tr>
<tr>
<td>Differences between radiologists at the first reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiologist A v B</td>
<td>0-40</td>
<td>5</td>
<td>30</td>
<td>9.0 (9.6)</td>
</tr>
<tr>
<td>Radiologist B v C</td>
<td>0-45</td>
<td>7</td>
<td>34</td>
<td>9.7 (10.2)</td>
</tr>
<tr>
<td>Radiologist A v C</td>
<td>0-63</td>
<td>5</td>
<td>34</td>
<td>9.9 (11.4)</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td>9.5 (10.4)</td>
</tr>
</tbody>
</table>

Discussion

HOW SHOULD OBSERVER VARIATION BE MEASURED?

In this study, measured conventionally with kappa statistics, observer agreement was good to excellent and absolute differences, both between and within readers, were small. The results compared favourably with previous studies specifically considering variability in the reporting of carotid angiograms. O’Leary et al found mean absolute differences for within reader agreement of 8.1% and between reader agreement of 9.1% in a comparison of both conventional angiography and ultrasound measurement variability. Chikos et al, in a study assessing the reporting of unsubtracted conventional angiograms, found corresponding results of 6.04% for intraobserver and 8.64% for interobserver variability (mean results for three radiologists). Both of these studies used caliper measurements, and both used the correlation coefficient to indicate good agreement. Correlation is an inappropriate tool for assessing agreement in measurement comparison studies (although, if the residual deviation, routinely derived when calculating the correlation coefficient, can be used as an index of disagreement). Bland and Altman have suggested an alternative approach based on the SD of the differences to give 95% limits of agreement, together with plots of the differences against the means (the 95% limits corresponding roughly to the mean difference ± 2 SD). Such an analysis assumes that the level of agreement does not vary across the range of possible values, which may not be true in the case of carotid artery stenosis. However, the plots themselves do show how the differences between observers vary according to the severity of the stenosis being measured. For instance, it can be seen that radiologist A consistently reads mild/moderate degrees of stenosis as less tightly stenosed than radiologists B and C, whereas the reverse is true for tight stenoses. An alternative approach is to group the data into discrete categories and calculate the proportion of vessels that are assigned to the same category by each observer, using a measure such as the kappa statistic. A drawback to this approach is that the results obtained will depend not only on the level of agreement but also on how many categories and which cut off values are chosen. Also very small differences can result in disagreement if they occur close to the cut off value whereas considerable differences can be present within the same category. However, clinical decisions do have to be made and, provided cut off values reflecting current practice are chosen, this form of analysis can give useful information.

FACTORS INFLUENCING OBSERVER VARIATION

Variability in reporting angiograms is inevitable, even though the assessment of stenosis can be based on what seem to be definite, objective measurements, such as the diameter of the minimal residual lumen and the diameter of the distal internal, or distal common carotid arteries. This is not just because of errors that occur in using calipers but also because factors such as the selection of the “best” viewing angle, the part of the distal internal carotid or distal common carotid to measure, or even the point of maximum stenosis, remain subjective judgements.

The plots of the differences against the means suggest that in general, agreement was better for more severe stenosis, a similar finding to some previous studies. One possible explanation is related to the method of measurement. When estimating the diameter of the residual lumen, the same absolute error in measurement will have more impact on the resultant degree of stenosis if the degree of stenosis is mild. For example, consider a vessel with residual lumen diameter 9 mm and original lumen diameter estimated as 10 mm by one observer and as 11 mm by another. This calculated percentage stenosis will be 10% and
18.2% respectively, with a difference between observers of 8.2%. Now consider a vessel with residual lumen 1 mm and original lumen, as before, estimated as 10 mm by one observer and 11 mm by another. The resultant percentage stenosis measured by each observer is 90% and 90.9% respectively, a difference of only 0.9%. The same absolute error in estimating the position of the original lumen has resulted in nine times the disagreement when measuring mild stenosis.

IMPACT OF OBSERVER VARIATION ON CLINICAL PRACTICE

The introduction of intra-arterial digital subtraction angiography has generally been considered to be an improvement on conventional angiography. Smaller doses of contrast are required and subtraction films are immediately available. The sequence of films can be reviewed immediately while the patient is still on the table and magnified films can be prepared rapidly. All of these may lead to improvement in image quality, in interpretation, and, because of the lower doses of contrast, hopefully a lower risk to the patient. The digital subtraction technique can introduce structured noise into the system, which could impact on measurements of arterial calibre. In situations where extremely precise measurements are required, such as in the measurement of blood flow waveforms, techniques are available to minimise this problem. Even with improved image quality the human factor may still prevail and a certain degree of intraobserver variation is inevitable. The data we collected in this study allowed us to determine, in the light of the results of the European Carotid Surgery Trial, what impact such variation might have on clinical practice.

An important question is to what extent does observer variation in reporting affect the clinical decisions about surgery? The frequency with which the recommendation might change from “recommend surgery” to “avoid surgery” can be estimated by applying the criteria derived from the interim results of the ECST and NASCET. Because the “ECST-method” of measurement was used in this study, we have used the ECST criteria. Both trials reported a benefit for surgery over medical treatment for those patients with severe stenosis (>70%) of the symptomatic internal carotid artery. In addition, the ECST showed that patients with less than 30% stenosis of the symptomatic artery should avoid surgery. The number of vessels in this study in which the clinical decision differed according to which radiologist reported the films or with the same radiologist reporting the same film on two separate occasions, ranged from 3.4% to 7.3%, or about one in 20 vessels. With a precise cut off value, such as 70%, the actual variation required to classify vessels differently need not be large and indeed, could occur with a difference as small as 1%.

As might be anticipated, in a small proportion of cases, intraobserver variation in the interpretation of angiograms led to some patients who might benefit from surgery being categorised as having a degree of stenosis in which surgery was either not indicated or of uncertain value. The converse was also true in that some patients in whom the value of surgery was unclear were perhaps erroneously classified as having a lesion in which surgery was definitely indicated.

IMPACT OF OBSERVER VARIATION ON METHOD COMPARISON STUDIES

An important implication of the variability in reporting conventional angiograms is in the assessment of alternative methods of measuring carotid artery stenosis, such as duplex ultrasound and magnetic resonance angiography. In these comparisons x-ray angiography is rightly used as the standard for comparison, but it must be appreciated that perfect agreement between any alternative method and conventional angiography will not be possible, given the variation inherent in reporting conventional angiograms. Observer variation will affect the decision about surgery whatever method is chosen to image the carotid artery.

CAN THE FINDINGS IN OUR CENTRE BE GENERALISED TO OTHERS?

If observer error occurs in a regional neurosciences unit with a specific interest in cerebrovascular disease, using state of the art equipment, and highly trained neuroradiologists working to the high specification of a research study, it is likely that important variation might be more frequent in other centres where other types of angiography are used, where films are not necessarily reported by trained neuroradiologists, and where pressure of clinical workload may allow less time to be devoted to the interpretation of routine angiograms than would be the case in research studies.

Quality control of angiography and its interpretation is therefore an issue which needs to be tackled by prospective audit studies examining not just the complication rate of angiography as suggested by others, but also the interpretation of the films. Audit on this scale does require substantial resources. However, this will be most important if we are to avoid inappropriate referrals for carotid surgery or inappropriately denying surgery to patients who might benefit from it. These recommendations apply whatever imaging modality is used to identify those patients with transient ischaemic attack and minor stroke who are most likely to benefit from carotid surgery.

This study was undertaken while PAGS was a lecturer and honorary senior registrar in the Department of Clinical Neurosciences, Walton Hospital, Liverpool. PAGS and JS are now supported by grants from the Medical Research Council. This study was carried out with the support of the Stroke Association.

We acknowledge the help of the staff of the Department of Neuroradiology, consultants, radiographers, and clerks who provided invaluable assistance in the performance of the angiograms, collection of the data, copying of films, and other secretarial and clerical tasks. This study would not have been possible without their help and patience.

1 European Carotid Surgery Trialists’ Collaborative Group.
MRC European Carotid Surgery Trial: interim results.
Observer variation in the interpretation of intra-arterial angiograms and the risk of inappropriate decisions about carotid endarterectomy

for symptomatic patients with severe (70–99%) or with mild (0–29%) carotid stenosis. Lancet 1991;337:1235–43.


Observer variation in the interpretation of intra-arterial angiograms and the risk of inappropriate decisions about carotid endarterectomy.

G R Young, P A Sandercock, J Slattery, P R Humphrey, E T Smith and L Brock

*J Neurol Neurosurg Psychiatry* 1996 60: 152-157
doi: 10.1136/jnnp.60.2.152

Updated information and services can be found at:
http://jnnp.bmj.com/content/60/2/152

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**