Prognostic Value and Reproducibility of Measurements of Carotid Stenosis
A Comparison of Three Methods on 1001 Angiograms

Peter M. Rothwell, MB, ChB, MRCP; Rod J. Gibson, MRCP, FRCR; Jim Slattery, MSc; Charles P. Warlow, MD, FRCP, for the European Carotid Surgery Trialists' Collaborative Group

Background and Purpose The use of three methods of measuring carotid stenosis, which produce different values on the same angiograms, has caused confusion and reduced the generalizability of the results of research. If the results of future studies are to be properly applied to clinical practice, and if noninvasive methods of imaging are to be properly validated against angiography, a single, standard method of measurement of stenosis on angiograms must be adopted. This method should be selected on the bases of its ability to predict risk of ipsilateral carotid distribution ischemic stroke and its reproducibility.

Methods The method of measurement of carotid stenosis used in the European Carotid Surgery Trial (ECST), that used in the North American Symptomatic Carotid Endarterectomy Trial (NASCET), and a method based on the measurement of the common carotid (CC) lumen diameter were studied. Their use in the prediction of ipsilateral carotid distribution ischemic stroke was assessed in 1001 consecutively selected patients randomly assigned to medical treatment in the ECST. Carotid stenosis was measured by two observers working independently, using all three methods, on the angiographic view that showed the most severe stenosis of the symptomatic carotid bifurcation. Interobserver agreement was determined, and 50 angiograms were remeasured to determine intraobserver agreement.

Results There was little difference in the ability of the three methods to predict ipsilateral carotid distribution ischemic stroke. The CC method was consistently the most reproducible of the three, particularly for stenosis in the clinically important range of 50% to 90%.

Conclusions The CC method of measurement should be adopted as the standard method of measuring the degree of carotid stenosis on angiograms. (Stroke. 1994;25:2440-2444.)

Key Words • angiography • carotid arteries • prognosis

The European Carotid Surgery Trial (ECST) and the North American Symptomatic Carotid Endarterectomy Trial (NASCET) showed that the degree of stenosis of the origin of the internal carotid artery is a major predictor of the risk of ipsilateral ischemic stroke, and consequently of the potential benefit from carotid endarterectomy, in patients with a recent transient ischemic attack or minor ischemic stroke.1,2 The decision whether or not to perform carotid surgery in these patients frequently hinges on a single measurement of stenosis.

Three methods of measuring stenosis on an angiogram have been proposed (Fig 1). All involve measuring the lumen diameter at the point of maximum stenosis, but the denominator in the equation used to calculate stenosis is different. The ECST method uses the estimated normal lumen diameter at the site of the lesion, based on a visual impression of where the normal arterial wall was before development of the stenosis. The NASCET method uses the diameter of a visible portion of disease-free internal carotid artery distal to the stenosis.3 The third method uses the diameter of the visible disease-free distal common carotid artery (CCA).4 The use of different denominators results in different measurements of stenosis on the same angiograms.5 Although by use of a linear equation it is possible to convert measurements made by each method to those made by the others,6 if the results of future research and clinical trials are to be properly applied to clinical practice a single, standard method should be adopted by all. This would increase the generalizability of research findings and would facilitate the validation of noninvasive methods of imaging, which will inevitably replace conventional angiography with the latter's attendant risks and costs.

How should a standard method be selected? The ability to use a method to predict the risk of ipsilateral ischemic stroke is probably the most important criterion. Which method is best suited for distinguishing between patients at high risk and low risk of stroke? The three methods are unlikely to vary significantly in this respect, but this remains to be proven. Therefore, our first aim was to measure the ability of each of the methods to be used to predict the risk of ipsilateral carotid distribution ischemic stroke.

The reproducibility of measurements made by each of the methods may differ. Reproducibility of measurements of angiographic vessel stenosis can be exceedingly poor.6,7 Poor reproducibility of measurement of carotid stenosis undermines selection of patients for endarterectomy and reduces the overall efficacy of the procedure. If the prognostic value of stenosis measured by each of the methods were similar, the most reproducible method would be the logical choice as a standard. Therefore, our second aim was to determine the interobserver and intraobserver variability of measurement of stenosis by each of the three methods on the same angiograms.
Reproducibility

Correlation coefficients are often used,9 14 but they do not measure agreement and ignore any bias between observers. To measure agreement, continuous variables are usually subdivided into a number of categories, and the agreement in assignment of a stenosis to a given category is assessed.

Methods

The details of the 1001 consecutively selected ECST carotid angiograms studied are reported in the companion paper to this one.7 No angiograms were excluded. In brief, they comprised 789 selective arteriographic angiograms, of which 307 were digitally subtracted; 174 aortic arch angiograms, of which 92 were digitally subtracted; 29 intravenous digital subtraction angiograms; and 9 angiograms in which the technique was not clear. The mean age of the patients studied was 62.1 years (SD, 7.8), and 71% were male.

Two observers (P.M.R. and R.J.G.), working independently, measured the degree of stenosis of the symptomatic carotid artery using each of the methods detailed in Fig 1. In patients with bilateral symptoms, the most stenosed artery was measured. Measurements were made using a jeweller’s eyepiece graduated in tenths of millimeters on the single angiographic view that showed the greatest stenosis. The same measurement of the minimum residual lumen diameter was used for all of the methods. Each observer was blind to measurements made by the other and to the clinical details. No marks indicating the points of measurement were made on the angiogram films by either observer. To determine intraobserver variation, each observer repeated the measurement of 50 randomly selected angiograms 6 months after the first reading.

The outcome used to assess the prognostic value of the methods of measurement of stenosis was carotid distribution ischemic stroke, lasting longer than 7 days, ipsilateral to the measured stenosis. Mean follow-up was 5 years (range, 4 months to 12 years). Further definitions and details of follow-up have been published previously.1

Statistical Analysis

Prognostic Value

Analysis of the prognostic value of measurements made by each of the three methods was based on the area under the receiver operating curve derived from the predictive properties of the mean of the measurements made by the two observers. For each method, the 1001 mean measurements were ranked according to degree of stenosis and divided into deciles. The sensitivity and specificity of the three methods above and below each decile were determined using Kaplan-Meier estimates of 3-year stroke risk, calculated as described previously.8

Reproducibility

No single statistic is appropriate for summarizing reproducibility of measurement of a continuous variable such as carotid stenosis. Correlation coefficients are often used,9 14 but they do not measure agreement and ignore any bias between observers. To measure agreement, continuous variables are usually subdivided into a number of categories, and the agreement in assignment of a stenosis to a given category is assessed.

However, the results obtained depend on the number and size of the categories chosen. Few published studies of measurement of carotid stenosis have used comparable categories.12 Moreover, even if identical categories are chosen, agreement will vary depending on the distribution of measurements across the categories. Reproducibility of measurement of stenosis is also likely to vary with the degree of stenosis, and this variation cannot be described by a single statistic. For these reasons use of the $\kappa$ statistic13 and the intraclass correlation coefficient14 is inappropriate.

To determine the relation between reproducibility and degree of stenosis we used the now well-established methodology of Altman and Bland.15 No variance-stabilizing transformation appears to be obviously applicable; hence the imprecision in measurement is analyzed within each decile of stenosis, a procedure that gives readily understandable results and can be represented graphically. Stenoses are designated as being in a particular decile according to the mean of the measurements made by the two observers. Imprecision in measurement within each decile is represented by the standard deviation of the perpendicular distances of all points from the diagonal that represents perfect agreement. To compare the precision of the three methods, the results of the different methods of measurement were transformed onto the ECST scale as described in the companion paper to this one.6

Calculation of measurement imprecision in this way eliminates any element of disagreement that may be due to systematic bias between the two observers. Doing this is worthwhile because systematic bias can be corrected, whereas little can be done about a measurement that is inherently imprecise. Moreover, the imprecision between two observers’ measurements is a property of the measurements as well as of the observers and may be generalizable, whereas the bias is solely a property of the observers and is not generalizable. To calculate bias, the mean of each pair of measurements was used to classify the stenosis as being in a particular decile. The bias within a decile is the mean of the differences between the two measurements for all stenoses in the decile.

Data are also presented in the more traditional form as the percentage agreement and as the $\kappa$ statistic for agreement between the observers in the categorization of stenosis as mild (0% to 29%), moderate (30% to 69%), or severe (70% to 99%). Intraclass agreement was calculated by combining the measurements made by both observers and comparing the first measurements with the second measurements.

Results

Prediction of Ipsilateral Ischemic Stroke

The receiver operating curves representing the predictive power of each of the three methods of measurement were almost identical (Fig 2), indicating no significant difference between the predictive powers of the methods.

Reproducibility

For each of the methods the measurements of one observer are plotted against those of the other in Fig 3. The spread of measurements provides a visual measure of the extent of disagreement. Subjectively, the spread is least with the CC method and most with the ECST method. More objectively, the $\kappa$ statistic (SE) for interobserver agreement in the categorization of stenoses as mild, moderate, or severe was 0.76±0.02 for the common carotid (CC) method, 0.72±0.02 for the NASCET method, and 0.66±0.02 for the ECST method. The difference between the ECST and NASCET methods was not significant. The CC method was significantly better than the ECST method but not the NASCET method. The same trend was reflected in the percent disagreement in the classification of stenoses into the same categories.

Fig 1. Measurement of carotid stenosis on an angiogram: the equation used in the European Carotid Surgery Trial (ECST), that used in the North American Symptomatic Carotid Endarterectomy Trial (NASCET), and that in which the diameter of the common carotid (CC) artery is used as the denominator.

### Statistical Analysis

#### Prognostic Value

Analysis of the prognostic value of measurements made by each of the three methods was based on the area under the receiver operating curve derived from the predictive properties of the mean of the measurements made by the two observers. For each method, the 1001 mean measurements were ranked according to degree of stenosis and divided into deciles. The sensitivity and specificity of the three methods above and below each decile were determined using Kaplan-Meier estimates of 3-year stroke risk, calculated as described previously.8

#### Reproducibility

No single statistic is appropriate for summarizing reproducibility of measurement of a continuous variable such as carotid stenosis. Correlation coefficients are often used,9 14 but they do not measure agreement and ignore any bias between observers. To measure agreement, continuous variables are usually subdivided into a number of categories, and the agreement in assignment of a stenosis to a given category is assessed.

However, the results obtained depend on the number and size of the categories chosen. Few published studies of measurement of carotid stenosis have used comparable categories.12 Moreover, even if identical categories are chosen, agreement will vary depending on the distribution of measurements across the categories. Reproducibility of measurement of stenosis is also likely to vary with the degree of stenosis, and this variation cannot be described by a single statistic. For these reasons use of the $\kappa$ statistic13 and the intraclass correlation coefficient14 is inappropriate.

To determine the relation between reproducibility and degree of stenosis we used the now well-established methodology of Altman and Bland.15 No variance-stabilizing transformation appears to be obviously applicable; hence the imprecision in measurement is analyzed within each decile of stenosis, a procedure that gives readily understandable results and can be represented graphically. Stenoses are designated as being in a particular decile according to the mean of the measurements made by the two observers. Imprecision in measurement within each decile is represented by the standard deviation of the perpendicular distances of all points from the diagonal that represents perfect agreement. To compare the precision of the three methods, the results of the different methods of measurement were transformed onto the ECST scale as described in the companion paper to this one.6

Calculation of measurement imprecision in this way eliminates any element of disagreement that may be due to systematic bias between the two observers. Doing this is worthwhile because systematic bias can be corrected, whereas little can be done about a measurement that is inherently imprecise. Moreover, the imprecision between two observers’ measurements is a property of the measurements as well as of the observers and may be generalizable, whereas the bias is solely a property of the observers and is not generalizable. To calculate bias, the mean of each pair of measurements was used to classify the stenosis as being in a particular decile. The bias within a decile is the mean of the differences between the two measurements for all stenoses in the decile.

Data are also presented in the more traditional form as the percentage agreement and as the $\kappa$ statistic for agreement between the observers in the categorization of stenosis as mild (0% to 29%), moderate (30% to 69%), or severe (70% to 99%). Intraclass agreement was calculated by combining the measurements made by both observers and comparing the first measurements with the second measurements.

#### Results

#### Prediction of Ipsilateral Ischemic Stroke

The receiver operating curves representing the predictive power of each of the three methods of measurement were almost identical (Fig 2), indicating no significant difference between the predictive powers of the methods.

#### Reproducibility

For each of the methods the measurements of one observer are plotted against those of the other in Fig 3. The spread of measurements provides a visual measure of the extent of disagreement. Subjectively, the spread is least with the CC method and most with the ECST method. More objectively, the $\kappa$ statistic (SE) for interobserver agreement in the categorization of stenoses as mild, moderate, or severe was 0.76±0.02 for the common carotid (CC) method, 0.72±0.02 for the NASCET method, and 0.66±0.02 for the ECST method. The difference between the ECST and NASCET methods was not significant. The CC method was significantly better than the ECST method but not the NASCET method. The same trend was reflected in the percent disagreement in the classification of stenoses into the same categories.

---

Fig 1. Measurement of carotid stenosis on an angiogram: the equation used in the European Carotid Surgery Trial (ECST), that used in the North American Symptomatic Carotid Endarterectomy Trial (NASCET), and that in which the diameter of the common carotid (CC) artery is used as the denominator.
1 - Specificity

Fig 2. Receiver operating curves for the power of each of three methods of measuring stenosis to predict the Kaplan-Meier risk of ipsilateral carotid distribution ischemic stroke at 3 years: the European Carotid Surgery Trial (ECST) method (dashed line); the North American Symptomatic Carotid Endarterectomy Trial (NASCET) method (dotted line); and the common carotid (CC) method (solid line).

(Table 1) and the number of occasions on which the observers' measurements were different by less than 1% stenosis (CC method, 204 [20.4%]; NASCET method, 186 [18.6%]; ECST method, 160 [16.0%]). Table 2 shows that the CC method produced the highest level of agreement between observers by 5% stenosis or less and the lowest

TABLE 1. Comparison of Classification of Stenoses as Mild (0-29%), Moderate (30-69%), and Severe (70-99%) by Two Observers for Each of the Three Methods of Measuring Stenosis

<table>
<thead>
<tr>
<th>Observer A</th>
<th>Observer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild</td>
<td>ECST method*</td>
</tr>
<tr>
<td></td>
<td>Mild</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Severe</td>
</tr>
<tr>
<td>ECST method*</td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>94</td>
</tr>
<tr>
<td>Moderate</td>
<td>40</td>
</tr>
<tr>
<td>Severe</td>
<td>0</td>
</tr>
<tr>
<td>NASCET method*</td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>97</td>
</tr>
<tr>
<td>Moderate</td>
<td>40</td>
</tr>
<tr>
<td>Severe</td>
<td>0</td>
</tr>
<tr>
<td>CC method*</td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>113</td>
</tr>
<tr>
<td>Moderate</td>
<td>22</td>
</tr>
<tr>
<td>Severe</td>
<td>0</td>
</tr>
</tbody>
</table>

*Disagreement between Observer A and Observer B, 16%; †disagreement, 18%; and ‡disagreement, 20%.

Measurements using the NASCET method are transformed onto an ECST scale.
level of disagreement by more than 10% stenosis. The bias between the two observers was relatively small. For each method one observer’s estimations of stenosis were consistently higher than the other’s by 1% to 5%. Bias was greater for mild stenoses and decreased as stenosis increased. There was no overall difference between each of the three methods in terms of the bias between the two observers.

Imprecision of measurement was much greater than the bias for each of the methods, and fell steadily as stenosis increased. For stenoses of more than 80%, the imprecision of each of the methods was identical (Fig 4). For stenoses of less than 40% the ECST method was the most precise, and in the range of 50% to 90% the CC method was consistently the most precise.

Intraobserver agreement was greater than interobserver agreement. For the measurements by both observers combined the $\kappa$ value for intraobserver agreement in the categorization of stenoses as mild, moderate, or severe was 0.84 for the CC method, 0.78 for the NASCET method, and 0.68 for the ECST method ($SE = 0.06$ for all). The trend was identical when the measurements made by the observers were analyzed separately.

Discussion

That the three methods were of very similar prognostic value was expected and is consistent with the very similar results seen in the ECST and NASCET trials when identical ranges of stenosis were compared.\(^7\) The discussion will therefore concentrate on the reproducibility of the methods. However, it should be borne in mind that factors other than the degree of stenosis, such as plaque surface morphology or stenosis of the external carotid artery, may also be of prognostic value, and an index combining a number of factors may eventually prove to be most appropriate.

Considering the importance of the degree of carotid stenosis in the management of patients with cerebrovas-

cular disease, there has been remarkably little published research on the reproducibility of its measurement on angiograms. Chikos et al\(^7\) examined the observer variability of measurement of stenosis using the ECST method on the carotid angiograms of 100 consecutive patients. The clinical details and the indications for angiography were unclear, and 36 angiograms were excluded because they did not meet an unspecified quality standard. Moreover, the majority of angiograms measured were of vessels with only mild stenosis, so the results are of little relevance to present-day clinical practice. Brown and Johnston\(^10\) determined the observer variability of quantitative measurement of stenosis on selected high-quality arterial angiograms, but did not actually quantify the variability. Murie and McKay\(^13\) found interobserver and intraobserver agreements of 74% and 83%, respectively, for categorization of stenosis, measured using the NASCET method on 100 randomly selected angiograms, into six categories (0% stenosis, less than 25% stenosis, 25% to 49% stenosis, 50% to 75% stenosis, 75% to 99% stenosis, and occlusion). The reproducibility of the ECST and the NASCET methods of measurement of carotid stenosis have not been compared, and the reproducibility of the CC method has never been studied.

Judging by the overall percentage agreement and the $\kappa$ statistics for interobserver and intraobserver agreement in categorization of stenoses as mild, moderate, or severe, the CC method is the most reproducible. However, such single-figure assessments of overall agreement obscure any variation in agreement with the degree of stenosis, and cannot indicate which method has the highest level of agreement in the clinically important range of 50% to 90% stenosis (measured by the ECST method), in which significant variability might influence the decision to recommend surgery.

More information can be gained from the plots of one observer’s data against the other’s (Fig 3). For each of the three methods the variation in measurements decreased with increasing stenosis. The analysis of imprecision and bias by decile of stenosis quantifies this. The bias between observers in the measurement of stenosis was relatively small and contributed little to the overall disagreement. However, this result applies only to the two observers in this study and cannot be generalized; the bias between other observers may be greater. The imprecision of different observers’ measurements is more generalizable, and can be regarded as a measure of the disagreement expected between two observers.

![Graph of imprecision (defined as differences between two observers’ findings of percent stenosis) in the measurement of carotid stenosis, reported by decile of stenosis, calculated by the following methods: the European Carotid Surgery Trial (ECST) method (dashed line), the North American Symptomatic Carotid Endarterectomy Trial (NASCET) method (dotted line), and the common carotid (CC) method (solid line). NASCET measurements were transformed to the ECST scale.](image)

**TABLE 2. Measurements of Stenosis Made by Two Observers for Each of the Methods of Measurement Grouped According to the Degree to Which They Differed**

<table>
<thead>
<tr>
<th>Disagreement (% stenosis)</th>
<th>ECST</th>
<th>NASCET</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5</td>
<td>402</td>
<td>440</td>
<td>459</td>
</tr>
<tr>
<td>6-10</td>
<td>303</td>
<td>300</td>
<td>297</td>
</tr>
<tr>
<td>&gt;10</td>
<td>296</td>
<td>261</td>
<td>245</td>
</tr>
<tr>
<td>Total</td>
<td>1001</td>
<td>1001</td>
<td>1001</td>
</tr>
</tbody>
</table>

ECST indicates European Carotid Surgery Trial; NASCET, North American Symptomatic Carotid Endarterectomy Trial; and CC, common carotid.
who had no overall bias with respect to each other. It is therefore an approximate measure of the least disagreement likely between two observers. The imprecision was high in this study. For stenoses of 40% to 50% (measured by the ECST method), the imprecision of measurement between observers was approximately 10% for each of the methods. Disagreement will often be greater because imprecision is defined as the standard deviation of the overall range of differences within each decile. Measurements by the two observers often differed by more than 20% for mild or moderate stenoses (Fig 3). Disagreement between observers in the range of 50% to 90% stenosis (measured by the ECST method) is likely to be of greatest clinical importance. Within this range the imprecision was consistently least for the CC method.

This study has demonstrated that the three methods of measuring stenosis predict the risk of stroke equally well, but that the CC method is the most reproducible measure of the degree of stenosis on angiograms. The CC method is also likely to be the most easily measured by noninvasive imaging techniques. Given the significant risk of complications of carotid angiography,20,21 the transition to noninvasive imaging for selection of patients for endarterectomy seems inevitable. The CCA is more easily visualized using carotid duplex ultrasound techniques than is the internal carotid artery distal to the bulb, and by use of the CC method the difficulty in visualizing the bulb diameter in the presence of a calcified plaque would be avoided. The lack of turbulent flow in the CCA compared with areas distal to the bifurcation and stenosis is likely to result in better visualization on magnetic resonance angiography. Finally, the CCA is rarely so affected by disease that its normal diameter cannot be measured at some point. We therefore suggest that the CC method be adopted as a worldwide standard for the measurement of carotid stenosis on angiograms and as the basis of measurement of stenosis using noninvasive techniques. Future analyses of ECST data will include measurements of stenosis using the CC method. We hope that other researchers will follow suit.

Acknowledgments

Dr Rothwell's and Mr Slattery's research and the European Carotid Surgery Trial are funded by the United Kingdom Medical Research Council.

References


P M Rothwell, R J Gibson, J Slattery and C P Warlow

*Stroke*. 1994;25:2440-2444
doi: 10.1161/01.STR.25.12.2440

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1994 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

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/25/12/2440

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/