PURPOSE: To evaluate different carotid stenosis estimation methods with digital subtraction angiography. METHODS: We assessed the intraobserver reproducibility and interobserver variability of visual interpretation and the measurement methods used by the European Carotid Surgery Trial and the North American Symptomatic Carotid Endarterectomy Trial. Angiographic stenosis measurements according to both criteria were performed twice by a radiologist, a neurologist, and a vascular surgeon. Eighty bifurcations of consecutive symptomatic patients underwent 480 pairs of measurements. In addition, four radiologists estimated the stenoses visually.

RESULTS: Intraobserver consistency was slightly better by the European (κ, 0.86 to 0.94) than by the North American (κ, 0.68 to 0.91) trial criteria or by visual interpretation (κ, 0.79 to 0.81). No significant interobserver variability was found, except in the subgroup of mild stenoses by the North American Trial criteria. By κ statistic, the interobserver agreement was excellent by the European trial method (κ, 0.72 to 0.86), good by the North American trial method (κ, 0.59 to 0.77), and good to excellent by visual evaluation (κ, 0.68 to 0.88). The visual estimation agreed more closely with the European (κ, 0.73 to 0.92) than with the North American trial (κ, 0.35 to 0.74) criteria measurements. CONCLUSIONS: All three methods have good reproducibility in digital subtraction angiography. Interobserver differences become more important in the estimation of mild stenosis.

Index terms: Arteries, carotid; Arteries, stenosis and occlusion; Angiography, comparative studies; Brain, infarction


Several large randomized multicenter trials have been undertaken to evaluate the effect of carotid endarterectomy in the prevention of ischemic brain infarction on symptomatic and asymptomatic patients (1–5). Both the recent North American Symptomatic Carotid Endarterectomy Trial (NASCET) (1) and the European Carotid Surgery Trial (ECST) (2) have produced significant information about the beneficial effect of carotid endarterectomy on symptomatic patients with ipsilateral carotid artery stenoses of 70% to 99%. These results have been widely accepted in clinical practice. NASCET showed also that the more severe the stenosis over 70%, the more prominent the risk reduction after carotid endarterectomy. In addition, ECST demonstrated a lack of benefit in the surgical group for mild stenoses of less than 30%. Both trials continue to recruit patients who have carotid stenoses of 30% to 69% on their prerandomization angiograms to determine the level of stenosis at which the beneficial effects of carotid endarterectomy are greater than the surgical risk.

Although the results of these two large clinical trials are convincing, it must be emphasized that the definition of stenosis degree is based on different methods (6, 7). The intertechnique differences between the NASCET and ECST measurement methods are significant, both in se-
verely (8, 9) and mildly stenotic bifurcations (10). Because the ambiguities in stenosis degree are confusing, achieving uniformity is desirable. Because of the definite clinical results of NASCET, its measurement method recently has been strongly advocated (6, 7); however, before any particular method is chosen, the interobserver variability and intraobserver reproducibility of the method should be assessed in all degrees of stenosis.

Studies on the reproducibility of these two measurement methods are scant. Some of them have been limited to the group of severely and moderately stenosed bifurcations (1, 5, 9), but in routine clinical work, the majority of the bifurcations evaluated by angiography are mildly stenosed. Chicos et al have evaluated observer variability in conventional angiography (11), but since then, intraarterial digital subtraction angiography has widely replaced conventional angiography (12, 13). Consistency of the carotid stenosis estimations therefore should be reevaluated by digital subtraction angiography (14).

The present study was undertaken to assess the reproducibility of these two most widely used measurement methods in an unselected group of symptomatic patients undergoing intraarterial digital subtraction angiography, with respect to the current threshold stenosis degrees of 30% and 70%, and to clarify the reasons for interobserver differences. In addition, we wanted to evaluate the value of visual interpretation. Even if visual interpretation is considered scientifically archaic (9), it is still probably the most widely used method in clinical practice (8).

### Patients and Methods

**Patients**

The study group consisted of 41 consecutive patients, who were referred by a neurologist for aortic arch and carotid artery angiography. The indications for angiography were hemispheric or retinal transient ischemic attacks in 20 cases, nondisabling strokes in 10 cases, both transient ischemic attacks and strokes in 5 cases, and other symptoms (vertigo, tinnitus, headache, and dysphasia) in 6 cases. Three patients had undergone previous carotid endarterectomy. Thirty-four of the patients were men and 7 women; mean age was 58 years (range, 34 to 72 years). Clinical auscultation revealed carotid artery bruits unilaterally in 8 cases and bilaterally in 4 cases.

**Angiography**

Intraarterial digital subtraction angiography studies were performed using 300 mg/mL of iohexol (Omnipaque, Nycomed A/S, Oslo, Norway) as the contrast material. In 38 examinations the common carotid arteries were cannulated selectively with a contrast material volume of 6 to 8 mL. Three examinations were performed with only unselective aortic arch injections using 25 to 30 mL of contrast material volume per injection. Both carotid bifurcations were studied in a minimum of two projections; in most cases three projections were used. The images were displayed and processed on a monitor with a 1024 x 1024 matrix, and representative images were recorded on film by a laser printer.

**Film Evaluation**

All films, without patient identification data or clinical information, were retrospectively read independently by a neuroradiologist (M.P.) and three general radiologists (R.V., H.M., and K.P.). All reviewers were asked to estimate visually the degree of stenosis in each common and internal carotid artery according to their normal clinical practice. The carotid stenoses were classified into the following four groups: (a) normal or less than 30%, (b) 30% to 69%, (c) 70% to 99%, and (d) occluded. The carotid bifurcation was excluded from the study if any of the reviewers considered the technical quality of the images nondiagnostic.

In the second phase, the degree of the stenosis in the internal carotid artery (or common carotid artery, if more significant) was measured according to the criteria described by the NASCET collaborators (Fig 1B) (1, 6) and the ECST collaborative group (Fig 1C) (2, 15). Both methods use the angiographic view that shows the greatest degree of narrowing. Both measurements were independently taken by a radiologist (R.V.), a neurologist (K.K.), and a vascular surgeon (H.T.) twice with at least 3 weeks' interval between the separate sessions. The results were given as whole percent units. The radiologist selected the image of each bifurcation to be measured by all three reviewers considered the technical quality of the images nondiagnostic.

For further analysis, a transparent film was placed over the angiographic image, and the contours of the residual lumen were drawn on it. After that, the imaginary line of the normal carotid bulb was extrapolated between the proximal and distal margins of the stenotic lesion (15) and drawn on the transparency. The diameter of the stenosed segment and the two reference diameters, perpendicular
to the vessel axis, were also marked. All measurerers used their own separate transparencies.

**Statistics**

For statistical analysis, the measured stenosis percents were divided into four clinically relevant classes: (a) mild (0% to 29%), (b) moderate (30% to 69%), (c) severe (70% to 99%), and (d) occluded. This classification has been used in the following results, unless otherwise mentioned. Pearson correlations between the individual measurerers were calculated. To test reproducibility of the measurements, the Kruskal-Wallis one-way analysis of variance for interobserver variability and $\kappa$ statistics for both intraobserver and interobserver variability were used. The upper limit of the $\kappa$ coefficient (+1) occurs only in the case of a perfect agreement. A $\kappa$ coefficient greater than 0.75 represents an excellent agreement, 0.40 to 0.75 a fair to good agreement, and less than 0.40 a poor agreement (16, 17). All data were analyzed using the SPSS/PC + 4.01 (SPSS, Chicago, Ill) and BMDP (BMDP Statistical Software, Los Angeles, Calif) statistical packages. The differences were considered statistically significant if $P < .05$.

**Results**

The films of two carotid bifurcations were considered nondiagnostic and were excluded from the study. Thus, altogether 80 bifurcations were evaluated. When further analyses for subgroups of bifurcations were performed, the arbitrary division into groups was done according to the average NASCET stenosis, which was calculated from both the measurements of all three measurerers. Using this standard, 57 bifurcations were classified as mildly, 6 as moderately, and 11 as severely stenosed. Six internal carotid arteries were occluded, including one bifurcation previously operated on.

**Interobserver Variability**

No significant interobserver variability was found in either the ECST or the NASCET ratings (Kruskal-Wallis one-way analysis of variance). The correlation coefficients between individual observers varied from 0.911 to 0.949 for the ECST ratings and from 0.882 to 0.963 for the NASCET ratings (0% and 100% stenoses were left out of the analysis). However, when the interobserver variability was calculated separately for the subgroups (mildly, moderately, and severely stenosed bifurcations), in the group of mild stenoses a significant variance between the NASCET ratings was found ($P < .001$), one observer (K.K.) having higher ratings than the two others. This wider variability by the NASCET method is evident in Figure 2, which shows graphically the distribution of the stenosis degrees by individual measurerers.

The $\kappa$ statistic for the interobserver agreement ranged from 0.72 to 0.86 (mean, 0.79) for the ECST measurements and from 0.59 to 0.77 (mean, 0.69) for the NASCET measurements, respectively. There was no significant interobserver variability in the visually esti-
mated stenoses (Kruskal-Wallis one-way analysis of variance). The $\kappa$ statistic for the interobserver agreement of the visual estimation ranged from 0.68 to 0.88 (mean, 0.76; Table 1). Because total occlusions are most easily uniformly recognized, they were left out of the data in the second analysis, which gave interobserver $\kappa$ values of 0.63 to 0.85 for the visual interpretation.

**Intraobserver Reproducibility**

Kappa statistics for intraobserver reproducibility by both measurement methods and visual interpretation (tested by readers R.V. and H.M.) are given in Table 2. The reproducibility of all three methods was good, the intraobserver variation of the ECST method being slightly smaller than that of the other two. When the occluded vessels were left out of the analysis, the intraobserver $\kappa$ values for visual estimation ranged from 0.75 to 0.78. The reproducibility of the methods was also checked by categorizing the patients into two groups: surgery recommended (70% to 99% stenosis) and others (0% to 69% stenosis or occluded). Then intraobserver $\kappa$ values were 0.75 to 0.94 for the NASCET method, 0.94 to 1.0 for the ECST method, and 0.86 to 0.94 for visual estimation. The mean differences between the two ECST ratings of each observer were 0.1% to 1.9% and between the two NASCET ratings 0.1% to 3.4%, respectively.

**Agreement between Visual Estimation and Measurements**

The $\kappa$ values for the intertechnique agreement between the visually estimated stenoses and the average ECST stenoses were slightly better than the corresponding coefficients between the visual estimation and the average NASCET stenoses for all four radiologists (Table 3). One observer (R.V.) evaluated the digital subtraction angiographic images by all three methods, the visual estimation being performed before the measurements. The $\kappa$ value of this measurerer was again better between the visual estimation and the ECST method (0.82) than between the visual estimation and the NASCET method (0.65).

**Reasons for Variability**

The most common reason for both interobserver and intraobserver differences in stenosis degree was the selection of the reference diameter. In the ECST method, it is possible...
TABLE 1: Kappa statistic for interobserver variability by visual interpretation (95% confidence limits)

<table>
<thead>
<tr>
<th>Observer</th>
<th>H.M.</th>
<th>K.P.</th>
<th>MP.</th>
</tr>
</thead>
<tbody>
<tr>
<td>R.V.</td>
<td>0.76 (0.64-0.89)</td>
<td>0.88 (0.78-0.97)</td>
<td>0.74 (0.61-0.87)</td>
</tr>
<tr>
<td>H.M.</td>
<td>0.73 (0.60-0.86)</td>
<td>0.68 (0.54-0.82)</td>
<td></td>
</tr>
<tr>
<td>K.P.</td>
<td>0.79 (0.67-0.90)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 2: Kappa statistic for intraobserver consistency by the ECST and NASCET measurement methods and visual interpretation (95% confidence limits)

<table>
<thead>
<tr>
<th>Observer</th>
<th>ECST Method</th>
<th>NASCET Method</th>
<th>Visual Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>R.V.</td>
<td>0.94 (0.88-1.0)</td>
<td>0.91 (0.82-1.0)</td>
<td>0.79 (0.67-0.91)</td>
</tr>
<tr>
<td>K.K.</td>
<td>0.88 (0.79-0.97)</td>
<td>0.76 (0.62-0.90)</td>
<td></td>
</tr>
<tr>
<td>H.T.</td>
<td>0.86 (0.76-0.96)</td>
<td>0.68 (0.53-0.83)</td>
<td></td>
</tr>
<tr>
<td>H.M.</td>
<td></td>
<td>0.81 (0.69-0.93)</td>
<td></td>
</tr>
</tbody>
</table>

to either exaggerate or underestimate the imaginary border of the normal bulb. In addition to the border with the most prominent plaque, most often the posterior wall, there were differences also in the tendency to extrapolate the opposite border depending on whether the plaque was considered circumferential. In the NASCET method, there was variation in the level at which the bulb was considered to end. Differences in the measured stenosed diameters were less common. In a few cases, different parts of a long stenosed segment were measured. There was also slight variation in the angles of the measurements, based on the different interpretation of the vessel axis.

Discussion

Atherosclerotic disease has a considerable clinical significance with its high prevalence and serious complications, mainly ischemic brain infarction. During the last few years, various treatment methods have been established to reduce the risk of brain infarction. At the same time, the diagnostic methods of carotid artery stenosis have undergone rapid development. Despite the advantages of new, noninvasive vascular imaging methods, such as color-coded Doppler ultrasound and magnetic resonance angiography, contrast angiography has remained the standard of reference in studies concerning carotid artery disease. Angiography, commonly digital subtraction angiography, also has been the standard of reference when the accuracy of the new imaging methods has been evaluated.

But how consistent is this standard itself? Alexandrov et al, in their recent study on duplex ultrasound, came to the conclusion that the NASCET and ECST methods of measurement consistently underestimate the "true" anatomic stenosis (8). The actual accuracy of digital subtraction angiography for grading carotid stenoses also has recently been questioned (14). In scientifically designed trials, the clinical proof of the benefit of endarterectomy is, however, based on linear measurements from arteriography (9) on both digital and conventional studies.

The NASCET collaborators checked the reliability of their assignments in 127 randomly selected bifurcations and found a high degree of interobserver consistency (κ, 0.89) (1). They categorized the stenoses into two groups (30% to 69% and 70% to 99%), which must be taken into account when comparing the κ values with those of the present study with four groups. The Veterans Affairs Cooperative Study Group also found a high degree of interobserver consistency (95%) in their multicenter trial of asymptomatic carotid stenosis (5). They used the NASCET measurement method, and the consistency was calculated with respect to the minimally acceptable stenosis rate of 50%. Although the reproducibility of the NASCET method has been proved good in the groups of patients with moderate and severe stenosis, these studies did not include patients with mild stenoses, who form the majority in the present study.

In our hands, the reproducibility of the ECST measurement method was better. It gave better interobserver and intraobserver agreements (Table 2) than the NASCET method. The visual estimation also correlated better with the ECST method (Table 3). Interestingly, the interobserver variability by the NASCET method was
significant in the subgroup of patients with mild stenoses.

It can be mathematically shown that equal changes in the measured reference diameter cause greater interobserver variation the milder the stenosis is. This was also obvious in our study. If the carotid bifurcation was mildly stenosed, the interobserver variability by the NASCET method was more prominent. The same phenomenon also has been evident when the intertechnique differences between the ECST and NASCET methods have been evaluated, the ECST method giving significantly a greater degree of stenosis (mean, 10.7 ± 8.2%) in the group of patients with mildly stenosed bifurcations (10). It is important to realize that the limited reproducibility of the method possibly could lead to a different treatment policy, especially in this subgroup of patients.

When the average NASCET and ECST stenoses of the same bifurcation were compared here, 15 bifurcations were classified as belonging to the no-surgery group (stenosis <30%) by the NASCET method but were candidates for randomization and possible surgery (stenosis 30% to 69%) by the ECST method, provided that the side correlated with the symptoms. One bifurcation definitely would have been recommended for surgery by the ECST measurement (stenosis ≥70%) but was considered moderately stenosed by the NASCET method (30% to 69%) and possibly could have been treated conservatively (10).

The reliability of the ECST method depends on the correct estimation of the normal diameter of the artery. This may be especially difficult if the atheroma is circumferential, and no penetrating ulcer reveals the presence of plaque. In unsubtracted films, as is the case in traditional film-changer angiography, possible calcifications in the arterial wall can be used as a guide (11), but calcifications are not uniformly shown in diseased vessels (18). In digital subtraction angiography, unsubtracted images are not routinely reviewed. Chikos et al measured the maximum percent diameter reduction, at 5% intervals, in a similar way as in the ECST study (11). The difference between the readers was 8.6 ± 9.5% and between the two readings of each reader 6.0 ± 8.1%. Our results gave mean differences of 0.1 to 1.9 ± 9.2% between the two ECST measurements, a comparable finding.

Many studies have demonstrated that subjective visual evaluation of an angiogram is inaccurate (19–21). A considerable amount of intraobserver and interobserver variability seems to exist even in the evaluation of stenoses in a phantom model of a straight vessel (19). The situation is even more complicated in the carotid bifurcation, in which, in addition to the division of the vessel, the convex shape of the normal bulb also must be taken into account. The amount of variation in the carotid shape between subjects and even between the right and left sides of the same subject must be considered. Yet, at many institutions, the degree of carotid disease is still estimated mainly visually in routine clinical work. This method also has been frequently used in comparative studies between two imaging methods. In the work of Litt et al, the visual estimation showed a 72% overall agreement between the observers, with the best agreement in the severe category (22). Our experience was that visual interpretation showed fair to good reproducibility. Careful measurements cannot, however, be abandoned. Our visual estimations correlated better with the ECST method.

In their study on duplex ultrasound, Alexandrov et al also found a closer correlation between “eyeballing” and the ECST measurements in the evaluation of angiograms (8).

We decided to evaluate the bifurcations without clinical information to avoid the possible influence on the stenosis estimation. One year after the first report from NASCET, Barnett et al expressed their concern that there seems to be a tendency to estimate upward to 70% from moderate stenosis (7). The degrees of threshold stenosis in the statistical analysis were chosen according to the recent results of the two symptomatic carotid stenosis trials (1, 2), and all bifurcations were analyzed together regardless of the correlation with the symptoms. The final clinical decision of course will depend on both stenosis degree and clinical data, neurologic symptoms as well as possible coexisting diseases. The role of carotid endarterectomy in the treatment of asymptomatic carotid stenosis remains to be defined (3, 23).

In conclusion, although the NASCET method has been strongly recommended, for measuring carotid stenosis, we can recommend the ECST method because of its good reproducibility, especially in mildly stenosed bifurcations. If an institution decides to change its previous mea-
measurement method to another, we recommend that both methods be used simultaneously for the transitional period to avoid confusion.

References

2. European Carotid Surgery Trialists’ Collaborative Group. MRC European carotid surgery trial: interim results for symptomatic patients with severe (70–99%) or with mild (0–29%) carotid stenosis. Lancet 1991;337:1235–1243