Critical Appraisal of the Design and Reporting of Studies of Imaging and Measurement of Carotid Stenosis

Peter M. Rothwell, MRCP; Sarah T. Pendlebury, MRCP; Joanna Wardlaw, FRCR; Charles P. Warlow, FRCP

Background and Purpose—Several hundred studies have been published over the last few years on imaging and measurement of carotid stenosis. Despite all this research, there is still no consensus about how best to image and measure stenosis. One possible explanation for this is that many of the studies have not been large enough or methodologically sound enough to allow useful conclusions to be drawn. We aimed to assess the design and methods of a random sample of published studies of imaging and measurement of carotid stenosis using 9 simple criteria.

Methods—A formal literature search was performed for studies of imaging and measurement of carotid stenosis. Two subsets were randomly selected for detailed assessment: 20 studies published before 1991 and 20 published between 1993 and 1997 (some years after the initial publication of the ECST and NASCET trials). The criteria used to assess the selected studies were as follows: prospective rather than retrospective study design; patient selection based on a consecutive series or a random sample; adequate detail of study population; adequate detail of imaging techniques; inclusion of all investigations, ie, patients with poor-quality imaging were not excluded; blinded assessment of images; adequate detail of derivation of measurement of stenosis from images or data; adequate data on the reproducibility of measurements of stenosis; and study powered according to a sample-size calculation.

Results—There were many basic methodological deficiencies in both subsets of studies, with relatively little evidence of improvement with time. For example, only 33% of studies were prospective, only 45% studied a consecutive or random selection of patients, and only 38% reported any data on the reproducibility of measurements. More than half of the studies satisfied ≤4 of the 9 quality criteria. However, there was considerable variation between studies, with 7 studies satisfying ≥7 criteria and 10 studies satisfying ≤2. No study was based on a sample-size calculation. The number of patients studied was often small, particularly in the more recent studies: median sample size was 100 in the 1970–1990 studies and 58 in the 1993–1997 studies (P<0.0001).

Conclusions—The design and reporting of published studies of imaging and measurement of carotid stenosis are poor and have not improved much in recent years. The majority of published studies are not of a sufficient standard to enable the results to be used to inform clinical practice. The utility of future studies could be improved considerably by better adherence to 9 simple methodological guidelines. (Stroke. 2000;31:1444-1450.)

Key Words: diagnostic imaging ■ measurement ■ carotid stenosis

The number of carotid endarterectomies being performed in the United States and Europe increased dramatically in the 1970s and early 1980s.1,2 There was a parallel increase in the frequency of carotid angiography1 and an increasing interest in noninvasive imaging. A literature search for studies of noninvasive methods of carotid imaging published between 1977 and 1993 found 568 relevant articles.3 Interest intensified after publication of the first results of the European Carotid Surgery Trial [ECST] and the North American Symptomatic Carotid Endarterectomy Trial [NASCET] in 1991,4,5 which showed that the degree of symptomatic carotid stenosis was a major predictor of risk of ipsilateral ischemic stroke for patients undergoing medical treatment and thus also an indicator of the benefit that could be derived from carotid endarterectomy. This led to a flurry of editorials and reviews on the different methods of measurement of stenosis6–14 and many additional studies comparing different imaging techniques. A manual search of Stroke from 1993 to 1998 (P.M.R.) revealed 42 such studies (excluding articles on vessel dissection or plaque morphology).

Carotid endarterectomy rates are rising again,15,16 and many clinicians are keen to use noninvasive methods, such as duplex ultrasound scanning, spiral CT scanning, and magnetic resonance angiography, to select patients for surgery. Conventional arterial angiography has significant morbidity and mortality,17,18 and it is now recognized that early retro-
spective studies probably underestimated these risks.\textsuperscript{19,20} A systematic review of the prospective studies of angiography in patients with cerebrovascular disease reported a risk of permanent neurological sequelae of 1\% and an overall mortality rate of 0.1\%,\textsuperscript{17} and significantly higher risks have been reported.\textsuperscript{19} Although the risks of arterial angiography may be lower in some very experienced centers, the procedure is still costly and time consuming, often requiring admission to hospital.\textsuperscript{21} For these reasons, it has become clear that we need, if possible, to progress to the routine use of noninvasive methods of carotid imaging in the selection of patients for endarterectomy.

The results of the clinical trials that demonstrated that endarterectomy was beneficial for symptomatic carotid stenosis were stratified by use of measurements of stenosis on arterial angiograms. There are no direct data to show that noninvasive methods of imaging can be used in the same way to differentiate between patients who should benefit from endarterectomy and those who may not. It is necessary, therefore, to validate noninvasive methods against conventional angiography. However, although many such studies have been published, the majority are undermined by poor design, inadequate sample sizes, and inappropriate analysis and presentation of data. It is at least partly as a consequence of this that there is still no consensus about how best to image the carotid artery. If noninvasive methods of imaging are to be properly validated and the findings of different studies compared, then a consistent and methodologically sound approach to study design and analysis must be adopted. This is particularly important given the large sample sizes required to accurately define the measurement characteristics of different imaging techniques and the consequent need for systematic reviews of studies of carotid imaging.\textsuperscript{3,22–24}

We assessed the design and reporting of a random sample of published studies of imaging and measurement of carotid stenosis using 9 simple criteria. Statistical analysis and presentation of results will be dealt with in a future paper. To assess whether the quality of studies has improved since the publication of the initial results of ECST and NASCET and the consequent realization of the importance of accurate measurement of the degree of carotid stenosis, articles published in 2 specific periods were studied: 1970–1990 (before ECST and NASCET) and 1993–1997 (some years after the publication of the initial results of ECST and NASCET).

**Methods**

**Inclusion Criteria**

We aimed to review articles that reported original data derived from imaging the carotid circulation in human subjects, the primary aim of which was either to determine the properties of the measurement of carotid stenosis using 1 imaging modality or to compare measurements of stenosis using 2 or more different imaging modalities. Studies could be confined to conventional angiography or a noninvasive method of imaging or could be a comparison of both. The review was confined to studies published in English-language journals.

**Search Strategy**

Articles were identified by the following means:

1. Articles already known to the authors from previous studies of the literature.
2. A single observer (P.M.R.) identified potential articles using CD ROM (Cambridge Medline 1970–1997). The search strategy was based on combinations of the following terms: carotid; stenosis; angiography; duplex; Doppler; ultrasound; ultrasonography; magnetic resonance imaging; magnetic resonance angiography; CT angiography.
3. The reference lists of all articles known to the authors and of articles identified from the literature were searched.

**Selection of Studies for Review**

Given the very large number of published studies, it was not possible to review all those identified in the literature search. Therefore, of the studies that fulfilled the inclusion criteria, 20 were randomly selected from each of the 2 publication periods. All of the articles published in each of the 2 periods were numbered sequentially (by S.T.P.). Another reviewer (P.M.R.) then used the random number–generating function in SPSS (version 7.0) to select 20 numbers at random from the range of numbers available for each period. Forty articles were thus selected for detailed study.

**Methodological Criteria**

Each article was assessed by 2 reviewers (P.M.R. and S.T.P.). The number of patients studied was recorded, and 9 methodological criteria were used to assess the quality of the articles (see below). Each of the criteria is considered in detail in the Discussion. If the 2 reviewers disagreed in their assessment of an article, they reassessed the article together until a consensus was reached. The criteria used to assess the selected studies were as follows:

1. Prospective rather than retrospective study design.
2. Patient selection based on a consecutive series or a random sample.
3. Adequate detail of study population (age, sex, clinical presentation, and indications for investigation).
4. Adequate detail of imaging techniques (sufficient for the study to be repeated).
5. Inclusion of all investigations, ie, patients with poor-quality imaging were not excluded.
7. Adequate detail of derivation of measurement of stenosis from images or data (sufficient for the study to be repeated).
8. Adequate data on the reproducibility of measurements of stenosis (data on either intraobserver agreement or interobserver agreement were accepted).
9. Study powered according to a sample-size calculation.

**Results**

The electronic literature search identified 486 publications that concerned carotid imaging. Exclusion of studies that did not fit our inclusion criteria left 132 studies. A further 13 studies that were already known to the authors and that fulfilled the inclusion criteria were not identified by the search. A search of the reference lists of these 145 studies revealed an additional 17 studies that fulfilled the inclusion criteria. Of the resulting sample of 162 studies, 93 were published before 1991 and 69 were published between 1993 and 1997. Twenty studies from each period were randomly selected for detailed review (1970–1991\textsuperscript{25–44} and 1993–1997\textsuperscript{45–64}). No studies by any of the authors were selected, and thus there was no conflict of interest.

The methodological assessments of the studies are shown in Table 1. There were many methodological deficiencies in both cohorts, with relatively little evidence of improvement with time. For example, only 33\% of studies were prospective, only 45\% studied a consecutive or random selection of patients, and only 38\% reported any data on the reproduc-
TABLE 1. Forty Randomly Selected Studies of Imaging of Carotid Stenosis Assessed According to 9 Simple Methodological Criteria

<table>
<thead>
<tr>
<th>Number of Studies Fulfilling Criteria (% and 95% CI)</th>
<th>1970–1990 Studies (n=20)</th>
<th>1993–1997 Studies (n=20)</th>
<th>Total (n=40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prospective design</td>
<td>4 (20%, 6–44)</td>
<td>9 (45%, 23–69)</td>
<td>13 (33%, 19–49)</td>
</tr>
<tr>
<td>Patient selection: consecutive series or random sample</td>
<td>8 (40%, 19–64)</td>
<td>10 (50%, 27–73)</td>
<td>18 (45%, 29–62)</td>
</tr>
<tr>
<td>Adequate detail of study population reported in article</td>
<td>3 (15%, 3–38)</td>
<td>15 (75%, 51–79)</td>
<td>18 (45%, 29–62)</td>
</tr>
<tr>
<td>Adequate detail of imaging techniques reported in article</td>
<td>19 (95%, 75–100)</td>
<td>19 (95%, 75–100)</td>
<td>38 (85%, 83–99)</td>
</tr>
<tr>
<td>Inclusion of all investigations, ie, patients with poor-quality imaging not excluded</td>
<td>7 (35%, 15–59)</td>
<td>3 (15%, 3–38)</td>
<td>10 (25%, 13–41)</td>
</tr>
<tr>
<td>Blinded assessment of images</td>
<td>11 (55%, 32–77)</td>
<td>12 (60%, 36–81)</td>
<td>23 (58%, 41–73)</td>
</tr>
<tr>
<td>Adequate detail of method of derivation of measurement of stenosis from images/data reported in article</td>
<td>12 (60%, 36–81)</td>
<td>18 (80%, 68–99)</td>
<td>30 (75%, 59–87)</td>
</tr>
<tr>
<td>Data reported on reproducibility of measurement of stenosis</td>
<td>6 (30%, 12–54)</td>
<td>9 (45%, 23–69)</td>
<td>15 (38%, 23–54)</td>
</tr>
<tr>
<td>Study powered according to a sample size calculation</td>
<td>0 (0%, 0–17)</td>
<td>0 (0%, 0–17)</td>
<td>0 (0%, 0–9)</td>
</tr>
</tbody>
</table>

No study was based on a sample-size calculation. Two studies looked at much large numbers of patients than the rest (500 patients40 and 400 patients63). However, the sample sizes in the remaining studies were often very small. This was particularly true of the more recent studies; median sample size was 100 in the 1970–1990 studies and 58 in the 1993–1997 studies (P<0.0001). Ten of the 1970–1990 studies had sample sizes of 100 or more compared with only 2 of the more recent studies.

Discussion

It is clearly important that effective investigations should be encouraged and ineffective investigations abandoned. The only way in which this will be achieved is by the performance, publication, and dissemination of high-quality and adequately powered clinical studies and systematic reviews. However, the design and reporting of the majority of studies reviewed here did not reach a sufficient standard. Although there is no gold standard with which to assess the quality of a study, and the choice of assessment criteria could be criticized, the 9 criteria could not be said to be unreasonable. The fact that some studies satisfied nearly all the criteria suggests that they were not unrealistic. Each of the methodological criteria that were used in this study is justified below.

Study Design (Criterion 1) and Selection of Patients (Criterion 2)

Studies should concentrate on patient populations that are comparable to patients seen in ordinary clinical practice. The answer to an important question is of little value if asked of the “wrong” patients. For example, a study of the reproducibility of measurement of stenosis on angiograms from a group of patients in which the majority had a carotid stenosis of <30% by the ECST method (<0 by the NASCET method)26 will tell us very little about measurement in patients with moderate and severe stenosis, the group in which variability might affect clinical decision making. Similarly, a study of intravenous digital subtraction angiography in predominantly young, fit patients will give much better results than an identical study in an older, more relevant population with widespread vascular disease; the causes of inadequate visualization of the stenosis with intravenous angiography, such as movement artifact, cardiac failure, and poor respiratory function, increase in frequency with age and concurrent disease. The population studied is of particular importance in studies of the complications of imaging procedures. Studies of the complications of conventional selective arterial angiography that concentrated on an elderly population with symptomatic cerebrovascular disease found much higher morbidity and mortality than studies in less-selected populations.18,19

Patients studied should be consecutive or random samples. Selection bias can undermine the generalizability of study results. For example, several of the 1970–1990 studies compared arterial angiography with intravenous digital subtraction angiography. However, most of these were retrospec-
TABLE 3. Forty Randomly Selected Studies of Imaging of Carotid Stenosis, Published During 2 Time Periods, Stratified According to the Number of Patients Studied

<table>
<thead>
<tr>
<th>Number of Patients Studied</th>
<th>1970–1990 Studies (n=20)</th>
<th>1993–1997 Studies (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–24</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>25–49</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>50–74</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>75–99</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>100–199</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>200–499</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>500+</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

If a patient has been put through an investigation and the result has been used to inform a clinical decision, it should be good enough to include in a study. Many of the 1970–1990 studies that compared the accuracy and reproducibility of measurement of stenosis using intravenous angiography with that using arterial angiography excluded intravenous investigations that were considered inadequate.30,34,40,41 Consequently, intravenous angiography compared well with conventional angiography. The results of these studies contrast with 2 studies that included all investigations, in which adequate views of both carotid bifurcations were seen in only 26%35 and 42%36 of patients imaged with intravenous angiography. Similarly, some studies of the reproducibility of measurement of stenosis on angiograms have been confined to angiograms selected on the basis of quality.25,26,29 Exclusion of poor-quality investigations undermines the relevance of these studies to clinical practice.

Criterion 6: Was the Assessment of Images Blinded to Other Information?
The need for blinding of observers to any information that might bias their measurements is self-evident. For example, in studies of the interobserver reproducibility of measurement of the degree of carotid stenosis with a particular technique, the observers should be independent and blinded to the findings of the others. Similarly, in studies comparing measurements made by different techniques, the observer should be blinded to any information that might lead to recognition of the fact that 2 investigations are from the same patient. Despite the importance of blinding, in nearly half the studies reviewed, it was neither implicit nor stated that observers were blinded.

Criterion 7: Was the Method of Measurement of Stenosis Described?
Although it is only in recent years that the disparities between measurements made by the different methods have been realized, it is still very surprising that 8 of the 1970–1990 studies did not define how stenosis had been measured on the angiograms.30–32,35,36,41,42,44 Of those studies that did give details, there was an even split between the ECST method25,28,29,34,40 and the NASCET method27,33,37,39,43 An exact definition of how the degree of stenosis is derived is particularly important in studies of noninvasive methods of imaging. Perhaps not surprisingly given the increased awareness of the disparities between the different methods of measurement, only 2 of the 1993–1997 studies failed to provide an adequate description of the method used.53,61 Of the remaining 1993–1997 studies, 10 used only the NASCET method,46,49,51,54–56,58,59,62 4 used only the ECST method,45,47,48,64 and 4 used both methods.52,57,60,62

Criterion 8: Were Data on the Reproducibility of Measurements Reported?
Interobserver agreement in the interpretation of radiological investigations may be little greater than that expected by chance alone.55–67 For example, reproducibility of measurements of stenosis on angiograms of the coronary or peripheral arterial circulations can be very poor.68,69 There is no reason
to assume that measurements of carotid stenosis by any technique will be less prone to observer variability. Any study that involves measurement of stenosis should provide some information about the reproducibility of the measurements made. Measurements are only likely to have clinical utility if they are reproducible. However, fewer than half of the studies reported any reproducibility data.

**Criterion 9: Sample Size**

Clinical trials often require very large sample sizes to measure the effectiveness of treatments with sufficient precision to influence clinical practice. The same principles should be applied to the validation of new methods of imaging. However, it is difficult to perform randomized controlled trials comparing different imaging methods. Trials would have to be vast to produce a reliable estimate of the effect of a method on eventual patient outcome. In practice, it is reasonable for imaging studies to compare a new technique with an established gold standard in the same group of patients and then extrapolate the results to estimate the likely outcome if the old technique were to be replaced by the new technique. However, although such studies will need much smaller sample sizes than randomized controlled trials that examine patient outcome, they still need to have the power to define any differences between the different techniques with clinically useful precision.

The sample size required will depend on the exact nature of the question being assessed. The simplest question, and the one that generally requires the smallest sample size, is assessment of the sensitivity and specificity of one test to detect a threshold defined by a gold standard, eg, 70% stenosis or complete occlusion as defined on conventional angiography. However, even this requires a large sample size to have clinically useful precision. For example, if we suppose that a population of 600 patients has a 20% prevalence of 70% to 99% carotid stenosis on conventional angiography (ie, 120 cases), and that carotid ultrasound correctly identifies 108 of these, the sensitivity of carotid ultrasound in the detection of severe stenosis would be 90%, but the lower 95% CI of this estimate is only 75%. In other words, even though the sample size was larger than any of the 40 studies reviewed here and was 10 times larger than many of them, the study still does not have the power to exclude clinically unacceptable false-negative rates.

None of the 40 studies reviewed were powered according to prespecified sample-size calculation. However, some of the small sample sizes in the 1970–1990 studies are understandable. Many of these studies were performed very early in the evolution of noninvasive methods, and the results were not intended to be applied directly to clinical practice. Moreover, the importance of accurate measurement of stenosis had not been demonstrated clearly. The situation was quite different, however, at the time of the 1993–1997 studies. Most of the imaging techniques studied were already used routinely and now required proper validation. This was implicit in many of the studies, several of which made clinical recommendations on the basis of their results. However, the sample sizes were completely inadequate in virtually all of the 1993–1997 studies. Indeed, on average, the later studies were much smaller than the earlier studies. A continuing profusion of small studies, many of which have inadequate methods, is likely to confuse rather than inform clinical practice.

**Meta-Analyses of Imaging Studies**

One way in which to extract some useful information from a group of small studies is to combine the results to increase precision. The techniques are now in place to allow useful meta-analysis of diagnostic studies. However, this is only possible if the design of studies and the analysis and presentation of data are of a sufficient and reasonably uniform standard. Our results suggest that this is not currently the case for published studies of imaging and measurement of carotid stenosis.

**Conclusions**

If the results of clinical trials of carotid endarterectomy are to be applied to clinical practice with noninvasive imaging, then new techniques must be properly validated against angiography. Review of previous research in this area shows that study methods and reporting of results are often poor. It will only be possible to apply the results of studies to clinical practice if the design, analysis, and reporting are of good quality. The quality standards set out in this article for study design provide a reasonable basis on which to proceed.

**References**


Critical Appraisal of the Design and Reporting of Studies of Imaging and Measurement of Carotid Stenosis
Peter M. Rothwell, Sarah T. Pendlebury, Joanna Wardlaw and Charles P. Warlow

*Stroke*. 2000;31:1444-1450

doi: 10.1161/01.STR.31.6.1444

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2000 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/31/6/1444

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/