Measurement Variability of Carotid Atherosclerosis: Real-Time (B-Mode) Ultrasonography and Angiography

Daniel H. O'Leary, MD, Fred A. Bryan, PhD, Marta W. Goodison, MPA, Matthew D. Rifkin, MD, Raymond Gramiak, MD, Marshall Ball, MD, M. Gene Bond, PhD, Rosalie A. Dunn, PhD, Barry B. Goldberg, MD, James F. Toole, MD, Hugh G. Wheeler, PhD, Nancy F. Gustafson, MS, Sven Ekholm, MD, and Jeffrey K. Raines, PhD

To quantify the within- and between-reader agreement of carotid B-mode ultrasonography and angiography interpretation, images from 117 patients examined by both modalities were read multiple times. Angiographic measurements were more reproducible than those of B-mode scans for all parameters except lesion width, but variations for B-mode scan measurements were similar to those for angiographic measurements. Within-reader agreement on the presence of ulceration was substantial for both modalities, whereas between-reader agreement was poor for B-mode scan and only moderate for angiography. (Stroke 1987;18:1011-1017)

Vascular ultrasonography is a minimally invasive technique, with no known risks, that can be used to evaluate individuals for clinically significant atherosclerosis and possibly to measure and track atherosclerosis in nonsymptomatic populations. To establish the capabilities of high-resolution B-mode ultrasound imaging (B scans) for detecting and quantifying atherosclerotic lesions in carotid and iliofemoral arteries of humans and animal models, a multicenter contract was issued by the National Heart, Lung, and Blood Institute (NHLBI). Five clinical centers received contract awards. An additional contract was awarded for a parallel study in primates and another for a data coordinating center. A 5-year research plan was begun in May 1981. The program was divided into 5 phases: protocol development (6 months), pilot studies (6 months), data collection including follow-up studies (36 months), and data analysis and manuscript preparation (12 months).

A Steering Committee composed of the principal investigators from each institution and scientific representatives from the NHLBI first met in June 1981 to begin developing formal hypotheses, common protocols, operation manuals, and criteria for patient eligibility. The following goals were established: 1) determination of the sensitivity, specificity, and accuracy of B scan with respect to angiography; 2) determination of the sensitivity, specificity, and accuracy of B scan and angiography with respect to pathologic specimen measurements; 3) evaluation of capabilities of B scan and angiography for detecting lesions of varying composition based on gross and microscopic examination of specimens; and 4) establishment of reproducibility in interpreting B scans and angiograms.

Subjects and Methods

Study Design

Two studies were initiated. The first was a sensitivity/specificity (S/S) study that was designed to compare results between ultrasonography, angiography, and pathology. The results of the S/S study are reported separately. The second was a repeatability/variability (R/V) study to evaluate within-reader variation (repeatability) and between-reader variation (variability) within a modality. This paper details the results of this second study, which was designed to determine the capabilities of different diagnostic methodologies. These results are based on multiple readings of single images, and thus the tested reproducibility is cross-sectional in time and not longitudinal. Since pathologic specimens were significantly altered during initial analysis, the repeatability procedures described here are limited to B scan and angiography of the carotid arteries.

After an initial 6 months of planning, a 6-month pilot study was undertaken, which was completed in June 1982. During this time, 119 patients were enrolled in the S/S and 26 in the R/V study. Subcommittees composed of angiographers, pathologists, statisti-
cians, and sonographers were convened to revise protocols based on lessons learned during this period. Final forms were developed for each of the following procedures: history and physical findings including lipid studies, B scan, angiography, functional studies, and surgical/pathologic findings. The pilot phase also allowed the sonographers to gain expertise and permitted the development of a uniform scanning technique. Quality control procedures were established for ultrasonography, including the routine use of in-house and circulating phantoms to ensure machine resolution over the course of the project.

Patient Selection

The main criterion for patient selection was carotid angiography for suspected cerebral vascular disease. Patients were excluded only if angiography was not performed or if the patient's clinical condition changed between the performance of the B scan and angiography. Patients were randomly assigned to S/S or R/V studies by the Data Coordinating Center, using prepackaged randomization envelopes and control sheets, within blocks to assure balance of assignments during patient recruitment. S/S patients were further randomized to treatments, reading of the B scan with (control) or without functional data. One patient was preselected for the R/V study from every block of 8–12 sequential patients at each clinical center. The block size varied with the anticipated enrollment of the center, minimizing the potential for selection bias due to factors such as time and experience.

On entry into the study, each patient had a history and physical examination form, including details of existing hypertension, cardiovascular disease, diabetes, cardiovascular symptoms or events, medications, peripheral vascular symptoms, smoking history, blood pressure, bruits, pulses, and physical characteristics, completed. These data are reported elsewhere. Among all R/V patients, the sonographer had no knowledge of the functional studies. Sonographers were permitted to employ but not record Doppler to distinguish between the internal carotid artery (ICA) and external carotid artery (ECA).

One half of the patients at each center were preassigned to each of the two sonographers. Ultrasound imaging was performed with 7.5- and 10-MHz transducers having an axial resolution of at least 0.5 mm. The right and left common carotid artery (CCA), ICA, and ECA were examined in transverse, anterolateral, oblique, and posteroblique planes. Images were recorded on videotape for later interpretation.

At least 2 views of the carotid bifurcation were routinely obtained at arteriography. Circular markers of known dimensions were placed on the skin adjacent to each vessel to permit correction for x-ray magnification. At the discretion of the angiographer, patients were examined by either conventional arterial injection and filming or by arterial or venous injection and digital subtraction filming.

Interpretation

Ultrasound and angiogram readers were unaware of history and physical findings. In one half of the S/S cases and in all the R/V cases, ultrasound readers had no knowledge of the functional studies. Ultrasound measurements were obtained from videotape images displayed on a television monitor frozen at peak systole. For each modality, the reader was required to choose the frame or film that they determined best demonstrated the minimum residual lumen (MRL). Measurements were made with calipers from this image to the nearest 0.5 mm, and the characteristics of any lesions noted were described. One of the two readers was randomly assigned to read the image again, for a total of 3 readings of each image. To prevent carryover knowledge, the R/V cases were interspersed among the much more numerous S/S studies and a minimum of 14 days was interposed between readings. The readers were blinded as to whether they were reading an S/S or R/V image.

Each reader first located the tip of the flow divider (Figure 1), which identified the origins of the ICA and ECA. A line drawn through the tip of the flow divider perpendicular to the axis of the ICA served as the boundary between the CCA and the ICA. The diameter of the CCA 15 mm proximal to the flow divider was measured and defined as the standard lumen because it is usually below significant disease. If the more distal carotid artery was thought to be normal, occluded, or

- Standard lumen size: 15 mm
- Reference distance: 2
- Lesion length: 4
- Lesion width: $4a + 4b$
- Minimum residual lumen: 5
inadequately delineated, the standard lumen was the only measurement obtained.

With a lesion, the reader specified the vessel segment containing the MRL. The plaque was designated as being on the near wall, far wall, near-far wall, or circumferential, and extension into adjacent segments of the carotid artery was noted. Plaque surface was characterized as smooth, irregular, or pocketed.

When a lesion was present, the distance from the flow divider to the point of MRL, termed the reference distance, was measured. Lesion width (LW), the combined thickness of the plaque on the near and far walls, was measured at the point of MRL. The site of maximal stenosis, was also measured.

Except for determination of standard lumen, measurements were reported only if, in the reader's judgment, the side was not occluded and a lesion was present on that side. If a reader judged a patient's B scan or angiogram to be substandard or unacceptable, the image was excluded from analysis. For both studies, images were judged unacceptable in those instances in which delimiters of MRL could not be observed on at least 1 longitudinal view. Analysis was limited to measurements made on diseased but patent arteries with the lesion having its MRL within 15 mm of the flow divider: MRL in mm, LW in mm, percent stenosis (defined as \((1 - \frac{MRL}{MRL + LW}) \times 100\)), reference distance in mm, and standard lumen in mm.

**Results**

Of the 1,099 patients enrolled in the carotid assessment program, 117 were randomly assigned to the R/V study. Excluded from analysis were 16% of the B scan and 9% of the angiograms judged by the readers to be substandard or otherwise unacceptable. Also excluded from analysis were the 26 R/V patients examined during the pilot phase of this study. Selective arterial filming was done in 74% and digital venous filming in 26% of the angiographic cases. These data were analyzed in combination and separately. Although a trend was observed for venous filming to be less reliable, no significant differences were found. Therefore, for the purposes of this report the results are combined, but these findings will be addressed in detail in a later publication. The maximum number of images available for analysis was 6 per patient, 3 each of the right and left carotid artery.

The unit of analysis was one side of a patient (artery). A patient's side was excluded from analysis if only 1 reading was available on that side or if all the reported readings on a side did not address the same arterial segment; the ICA and CCA were considered the same arterial segment for this purpose.

Furthermore, sides classified as normal or occluded on any reading were excluded from the R/V study analysis of MRL, LW, stenosis, and reference distance as no measurement other than standard lumen was recorded. The number of observations in the R/V study analysis varied little among measurements. Of the 117 patients assigned to the R/V study, 76 produced useful B scan measurements and 80 useful angiographic measurements. Nevertheless, the residual patients still provided between 150 and 160 sides for evaluation, and given the fact that multiple measurements were made on each side, the actual number of observations was 295 or more for determination of measurement error in each angiographic and 278 or more for each B scan measurement. Based on the standard error of the mean shown in Tables 4 and 5, these numbers are considered sufficient to estimate the measurement error within the two techniques, which was the primary objective of this evaluation.

**Data Analysis**

The statistical analysis determined the within- and between-reader agreement in B scan and angiographic measurement of lesion parameters. Three statistical measures of variability were used: the variance, the mean absolute difference between pairs of measurements, and the Pearson correlation coefficient between pairs of measurements.

The kappa statistic (κ), used to summarize the categorical data,\(^1\) is a standard measure of agreement in excess of chance between data classifications. In our study, the within-reader κ denotes the correlation among repeated independent classifications of an artery by the same reader. The between-reader κ denotes the correlation between independent (single) classifications of a single artery by different readers.

Following Landis and Koch,\(^4\) the following labels indicate the relative strength of agreement associated with the range of κ indicated in parentheses: poor (0–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.0). Although these labels are somewhat arbitrary, they do provide useful benchmarks.

**Categorical Data**

Within-reader agreement for lesion detection by B scan for all 3 carotid arteries and by angiography for the CCA and ICA was substantial (Table 1). Similarly, between-reader agreement about lesions of the CCA and ICA in both B scan and angiography was moderate and substantial, respectively, but the between-reader agreement varied little among measurements. Of the 117 patients assigned to the R/V study, 76 produced useful B scan measurements and 80 useful angiographic measurements. Nevertheless, the residual patients still provided between 150 and 160 sides for evaluation, and given the fact that multiple measurements were made on each side, the actual number of observations was 295 or more for determination of measurement error in each angiographic and 278 or more for each B scan measurement. Based on the standard error of the mean shown in Tables 4 and 5, these numbers are considered sufficient to estimate the measurement error within the two techniques, which was the primary objective of this evaluation.

**Data Analysis**

The statistical analysis determined the within- and between-reader agreement in B scan and angiographic measurement of lesion parameters. Three statistical measures of variability were used: the variance, the mean absolute difference between pairs of measurements, and the Pearson correlation coefficient between pairs of measurements.

The kappa statistic (κ), used to summarize the categorical data,\(^1\) is a standard measure of agreement in excess of chance between data classifications. In our study, the within-reader κ denotes the correlation among repeated independent classifications of an artery by the same reader. The between-reader κ denotes the correlation between independent (single) classifications of a single artery by different readers.

Following Landis and Koch,\(^4\) the following labels indicate the relative strength of agreement associated with the range of κ indicated in parentheses: poor (0–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.0). Although these labels are somewhat arbitrary, they do provide useful benchmarks.

**Categorical Data**

Within-reader agreement for lesion detection by B scan for all 3 carotid arteries and by angiography for the CCA and ICA was substantial (Table 1). Similarly, between-reader agreement about lesions of the CCA and ICA in both B scan and angiography was moderate and substantial, respectively, but the between-reader agreement varied little among measurements. Of the 117 patients assigned to the R/V study, 76 produced useful B scan measurements and 80 useful angiographic measurements. Nevertheless, the residual patients still provided between 150 and 160 sides for evaluation, and given the fact that multiple measurements were made on each side, the actual number of observations was 295 or more for determination of measurement error in each angiographic and 278 or more for each B scan measurement. Based on the standard error of the mean shown in Tables 4 and 5, these numbers are considered sufficient to estimate the measurement error within the two techniques, which was the primary objective of this evaluation.

**Data Analysis**

The statistical analysis determined the within- and between-reader agreement in B scan and angiographic measurement of lesion parameters. Three statistical measures of variability were used: the variance, the mean absolute difference between pairs of measurements, and the Pearson correlation coefficient between pairs of measurements.

The kappa statistic (κ), used to summarize the categorical data,\(^1\) is a standard measure of agreement in excess of chance between data classifications. In our study, the within-reader κ denotes the correlation among repeated independent classifications of an artery by the same reader. The between-reader κ denotes the correlation between independent (single) classifications of a single artery by different readers.

Following Landis and Koch,\(^4\) the following labels indicate the relative strength of agreement associated with the range of κ indicated in parentheses: poor (0–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.0). Although these labels are somewhat arbitrary, they do provide useful benchmarks.

<table>
<thead>
<tr>
<th>Arterial segment</th>
<th>Within-reader agreement (95% CI)</th>
<th>Between-reader agreement (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B scan</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCA</td>
<td>0.69 (0.58–0.80)</td>
<td>0.58 (0.46–0.71)</td>
</tr>
<tr>
<td>ICA</td>
<td>0.73 (0.62–0.83)</td>
<td>0.65 (0.53–0.77)</td>
</tr>
<tr>
<td>ECA</td>
<td>0.64 (0.49–0.79)</td>
<td>0.34 (0.15–0.52)</td>
</tr>
<tr>
<td><strong>Angiography</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCA</td>
<td>0.74 (0.64–0.83)</td>
<td>0.59 (0.48–0.71)</td>
</tr>
<tr>
<td>ICA</td>
<td>0.82 (0.72–0.92)</td>
<td>0.66 (0.54–0.79)</td>
</tr>
<tr>
<td>ECA</td>
<td>0.60 (0.47–0.73)</td>
<td>0.58 (0.45–0.70)</td>
</tr>
</tbody>
</table>

N, number of arteries; CI, confidence interval; B scan, B-mode ultrasonography; CCA, common carotid artery; ICA, internal carotid artery; ECA, external carotid artery.
agreement in the ECA was only fair in B scan and moderate in angiography.

For both B scan and angiography, within-reader agreement was substantial for the 0–49% and the 75% occluded stenosis categories, but was only moderate in the 50–74% category (Table 2). For angiography, between-reader agreement was also substantial in the 0–49% and 75% occluded categories and moderate in the 50–74% category. For B scan, however, between-reader agreement was substantial in the 75% occluded category, moderate in the 0–49% category, and only fair in the 50–74% category. For both within- and between-readers, overall agreement was substantial for angiography and moderate for B scan.

Table 3 presents the within- and between-reader agreement on the presence of ulceration. For this analysis, a lesion was classified as ulcerated by B scan if pocketing or irregularity was reported and by angiography if the presence of ulceration or an irregular lumen was reported. Within-reader \( \kappa \) for presence of ulceration were substantial for both B scan and angiography; between-reader agreement was poor for B scan and moderate for angiography.

**Measurement Data**

Within- and between-reader measures of variability are given in Table 4. The within-reader variance for B scan measurements of MRL, LW, reference distance, and standard lumen are larger than those for the corresponding angiographic measurements \( (p<0.05 \text{ for each}) \); the mean absolute differences show the same relations. The within-reader correlation for B scan measurements of MRL, stenosis, reference distance, and standard lumen are smaller than those for the corresponding angiographic measurements \( (p<0.05 \text{ for each}) \); the within-reader correlation for B scan measurement of LW is larger than that for the corresponding angiographic measurement.

The between-reader variances in the B scan measurements of all parameters except reference distance were significantly larger than those for the corresponding angiographic measurements \( (p<0.05 \text{ for each}) \); the same relation is reflected in the mean absolute differences for the two modalities. The between-reader correlations for B scan measurements of MRL, stenosis, and standard lumen are significantly smaller than those for the corresponding angiographic measurements \( (p<0.05 \text{ for each}) \).

Data were analyzed during the study at fixed time intervals to determine if a learning curve existed for either B scan or angiogram interpretation. None was detected.

**Discussion**

Because B-mode ultrasonography is used as a non-invasive method to diagnose carotid artery atherosclerosis and to follow plaque progression and regression over time, the reliability of ultrasound measurement must be known. Previous studies attempting to quantify this have used angiography as the standard for comparison, but our findings suggest that there are problems associated with both measurement processes. Although we have found that angiographic measurements are more reliable than B scan measurements for all parameters except LW, the variations within each modality are, in fact, similar.

Our results are consistent with other reports of observer variability in angiographic interpretation; to date, there have been no similar studies of B scan interpretation. In the largest study of carotid angiography reported, Chikos et al. found a similar degree of within- and between-reader variability in estimates of percent stenosis; their 64 cases were drawn from a
larger set of 100 consecutive patients, 36 of whom were excluded because of relatively poor technical quality. All studies were selective arterial injections; the single best opacified view from each projection was preselected, and each reader was given the same set of films. All measurements were made with calipers. Each angiogram was viewed twice by 3 readers. Their within-reader correlation for percent stenosis was 0.94 compared with our r of 0.85. Their average within-reader variance in estimating percent stenosis was 5.2% compared with our 8.1%. Their between-reader correlation was 0.92, and their mean absolute difference in percent stenosis was 7.2%; our corresponding figures were 0.85 and 9.1%. These differences are not large and are understandable in that theirs was a single-center study, their criteria for inclusion were more restrictive, and their readers were given no latitude in the selection of films to read.

Slot et al used κ to analyze between-reader agreement among 11 observers interpreting 21 single-plane translumbar arteriograms. κ ranged from 0.51 for the left superficial artery to 0.07 for the right profunda bifurcation. They concluded that, except for recognition of occlusion, so little agreement was found between observers as to raise serious doubt about the utility of angiography in clinical decision making. In a study of coronary angiography, Zir et al found a marked degree of between-observer variability in quantifying the percent coronary artery stenosis and noted that observer agreement decreased with increasing positive findings. They speculated that observer agreement would increase if the percent of normal vessels were higher or if only a single vessel were studied. They also focused on the difficulty of identifying normal arterial segments for determining percent stenosis in diseased vessels.

Optimum angiographic and B scan images still differ greatly in quality, primarily due to the physical principles upon which x-ray and acoustic transmission and deflection are based. Fundamentally, x-rays form sharper images because they scatter less than acoustic radiation. Theoretical resolution is also greater for x-rays because of the smaller wave lengths. Other considerations are the properties of biologic tissues that determine the transmission of x-rays and sound. Angiography produces arterial images by using contrast material to alter the transmission properties of the blood within the vessel. With this method, properties of the vessel wall are not altered, and thus the wall is not distinguished from surrounding tissue that has the same effect on x-rays. B scan imaging, on the other hand, differentiates surfaces of lesions and vessel walls by acoustic reflections that occur at the lumen–intima and lesion–tunica adventitia interfaces.

These different capabilities offer some clues regarding the cause of differences between angiographic and B scan measurements for lumen diameter and LW, both of which are higher for B scan. For angiographic images, LW must be calculated by comparing lumen diameter at the lesion location with the diameter at an unobstructed location. Since the base of atherosclerotic plaque is not seen, this calculation reveals only that part of the lesion which encroaches into the lumen, and whether the normal segment chosen is in fact normal is an assumption. For B scan images, LW is measured directly from the lesion–lumen to the lesion–tunica

### Table 4. Within- and Between-Reader Agreement

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Within-reader agreement</th>
<th>Between-reader agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variance (95% CI)</td>
<td>Mean absolute difference ±SEM (95% CI)</td>
</tr>
<tr>
<td>B scan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum residual lumen (mm)</td>
<td>0.76 (0.65–0.84)</td>
<td>1.00 (0.79–1.32)</td>
</tr>
<tr>
<td>Lesion width (mm)</td>
<td>0.81 (0.63–0.82)</td>
<td>1.48 (1.15–1.99)</td>
</tr>
<tr>
<td>Stenosis (%)</td>
<td>7.76 (6.60–6.84)</td>
<td>10.10 (7.94–13.42)</td>
</tr>
<tr>
<td>Reference distance (mm)</td>
<td>7.5 (5.7–10.6)</td>
<td>2.49 (0.33)</td>
</tr>
<tr>
<td>Standard lumen (mm)</td>
<td>0.44 (0.33–0.60)</td>
<td>0.64 (0.07)</td>
</tr>
</tbody>
</table>

**Angiography**

| Minimum residual lumen (mm)      | 0.17 (0.13–0.24)        | 0.42 (0.33–0.56)        | 0.6 (0.08)              | 0.85 (0.78–0.90)        |             |
| Lesion width (mm)                | 0.61 (0.60–0.81)        | 0.54 (0.43–0.70)        | 0.8 (0.07)              | 0.64 (0.50–0.75)        |             |
| Stenosis (%)                     | 8.1 (0.78–0.90)         | 82 (66–106)             | 9.1 (0.99)              | 0.85 (0.78–0.90)        |             |
| Reference distance (mm)          | 1.34 (0.22)             | 0.87 (0.80–0.91)        | 10.01 (7.82–13.40)      | 3.8 (0.39)              | 0.56 (0.40–0.68) |
| Standard lumen (mm)              | 0.48 (0.05)             | 0.89 (0.84–0.93)        | 0.29 (0.23–0.38)        | 0.5 (0.05)              | 0.83 (0.76–0.88) |

CI, confidence interval; B scan, B-mode ultrasonography.
adventitia interfaces, the sum of the lesion and arterial wall dimensions. Explaining why lumen diameter observed is greater by B scan than by angiography is more difficult. One hypothesis is that some arterial interfaces are delineated less clearly using B scans, particularly that the reflection from the intima-blood interface is weak. A stronger signal is usually reflected from the media-adventitia interface, and this interface may be incorrectly construed as defining lumen boundaries.

In this study ultrasound readers had a relative disadvantage because measurements were made by viewing a videotape and then freezing the image. There is a loss in image quality when comparing videotapes with real-time images. Also, it is extremely unlikely, considering the thousands of frames available, that repeat ultrasound readings were made from the same frame. Angiogram readers were presented with very few images, and the likelihood of choosing the same frame for repeat readings was high. How to compare B scan reproducibility was discussed often during the design of the study. On one hand, selection and distribution of still frames for interpretation would have been easier and more consistent with the procedure for interpreting angiograms. On the other hand, an important advantage of B scan imaging is that a dynamic display allows better appreciation of the wall-lesion-lumen interfaces. The decision to have readers view videotapes rather than still frames was based largely on providing readers with the best information possible at the expense of incurring greater variability since different sections of videotape would invariably be studied by different readers in many cases.

During development of the protocol, we attempted to develop a means by which to identify the precise location of a lesion within an arterial segment to compare modalities and to further identify specific arterial sites. For the carotid artery, the flow divider provided the best anatomic landmark. However, our data locating the most significant lesion with respect to the bifurcation by either modality have among the poorest reproducibility.

The explanation may lie in the fact that the CCA, ICA, and ECA do not usually lie in a single plane and, depending on the view, ECA and ICA may overlap for varying distances. The best view of the flow divider is often different from the best plane in which to view a lesion or MRL. Readers were required to make their measurements from that view which best demonstrated MRL, not that which best delineated the tip of the flow divider. In a study of observer variability in the measurement of coronary angiograms, Zir and his colleagues7 noted similar difficulty identifying reference locations and suggested this as a major source of error in the measurement process. Our finding has implications for longitudinal studies of atherosclerosis progression and regression since repeat measurements need to be made at the same site.

The reproducibility of detecting ulceration was much poorer than that for dimensional data, in part because precise criteria for designating smooth, irregular, or pocketed surfaces could not be developed, consistent with reports of other investigators.

Within- and between-reader agreement on the presence of ulceration differed markedly, particularly for B scan, suggesting that the detection of ulceration is quite subjective. Within-reader agreement measures consistency of detection; between-reader agreement measures the ability of a technique to detect ulceration on a consensus basis. Our findings indicate that each reader was able to establish criteria that he applied relatively consistently but that nevertheless were quite different from those used by other readers. Little consistency existed between readers on what constituted ulceration. Until better criteria are established for detecting ulceration, our results indicate that neither B scan nor angiography can produce a reliable diagnosis.

High-resolution B-mode ultrasonography has been suggested as a potential method for monitoring anatomic changes in atherosclerotic lesions over time. From our data, we calculate that, to be certain of disease progression at \( p < 0.05 \) and a power of 85% using the clinical methods described here, the minimum change in residual lumen for angiography is 2.7 mm and for ultrasonography is 3.2 mm. Because angiographic measurements are generally more precise than B scan measurements, the sample sizes needed for detecting changes in MRL are larger for B scan than for angiography. For example, the sample size needed for detecting (with \( p = 0.90 \) and \( \alpha = 0.05 \)) relative changes of the same magnitude by angiography will be about 15% smaller than that required for B scan. With a sample of 500 patients, one can detect a 4.3% change in MRL by angiography or a 5.9% change in MRL by B scan.

In clinical practice, the same reader and sonographer are more than likely to be involved in longitudinal follow-up of individual patients. Assuming no change in personnel, procedures, or instruments between evaluations, it is possible to detect a 9% (0.6 mm) change in standard lumen (with \( p = 0.90 \) and \( \alpha = 0.05 \)) by angiography but only a 17% (1.3 mm) change in standard lumen by B scan.

The safety and repeatability of B scan commends its wide use as a diagnostic screening procedure. We have demonstrated that the measurement variability involved in B-mode ultrasonography is similar to that in angiography. Our results suggest that B scan is an acceptable supplement to angiography and a reliable method to screen patients for carotid artery atherosclerosis. It appears to be a reasonable substitute for angiography to follow the time course of lesions in populations at risk for carotid atherosclerosis. However, if ultrasonography is to be used to quantitatively evaluate plaque progression and regression in arteries over time, our study emphasizes that the reproducibility of the technique should be established to ensure that apparent change is real.

References

1. Ricotta JJ, McKinney WM, Bryan FA, Bond MG, Kurtz A, O'Leary DH, Raines JK, Berson AS, Clouse ME, Calderon-


KEY WORDS • multicenter validation trial • repeatability • real-time B-mode ultrasonography • angiography • carotid artery.
Measurement variability of carotid atherosclerosis: real-time (B-mode) ultrasonography and angiography.
D H O'Leary, F A Bryan, M W Goodison, M D Rifkin, R Gramiak, M Ball, M G Bond, R A Dunn, B B Goldberg and J F Toole

Stroke. 1987;18:1011-1017
doi: 10.1161/01.STR.18.6.1011

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/18/6/1011

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/