An evaluation of the reproducibility of the measurement of the intima-media thickness of carotid arteries

Radosław Kaźmierski¹, Adam Niezgoda¹, Przemysław Guzik², Maria Łukasik¹, Wojciech Ambrosius¹, Wojciech Kozubski¹

¹Department of Neurology, Karol Marcinkowski University of Medical Sciences, Poznań, Poland
²Department of Cardiology-Intensive Therapy, Karol Marcinkowski University of Medical Sciences, Poznań, Poland

[Received 27 November 2002; Accepted 30 December 2002]

The intima-media thickness (IMT) of carotid arteries was demonstrated to be a reliable measure for early stages of atherosclerosis. B-mode ultrasound may be used to measure carotid IMT. The measurements of the IMT of the carotid artery (CA) conducted by different investigators can be comparable and enable the implementation of clinical trial successfully while maintaining a high reproducibility value. The objective of the study was to evaluate the reproducibility of the measurements made by the same investigator on two separate occasions (intraobserver variability) and the reproducibility of the off-line measurements between four sonographers in our laboratory (interobserver variability).

The IMT of CA in 25 subjects (15 post stroke and 10 healthy persons) was investigated with the use of high-resolution ultrasonography. The CA subdivided into the common, bulbs and internal segments were scanned twice with a 3-week interval. Additionally three other readers with different levels of experience and skills in ultrasonography were asked to perform the same measurements in duplicate with at least a 3-week interval between.

A high concurrence for intraobserver variability was detected with a correlation coefficient ranging from 0.92 to 0.95; p < 0.0001, and maximal bias 0.019 mm. Interobserver variability for all four readers also demonstrated a high correlation coefficient ranging from 0.72 to 0.83; p < 0.0001, and the maximal bias of measurements did not exceed 0.08 mm.

The analogue measurements performed by the team demonstrate a reliable reproducibility in terms of the results of morphologic measurements. The differences obtained in the study were less than the error of the method (i.e. 0.1 mm) and should not influence clinical decision-making. Additionally, this study demonstrated that interobserver concurrence increases with the increasing experience of the investigators.

key words: atherosclerosis, ultrasonography, interobserver, intraobserver, variability
INTRODUCTION

A normal large artery has a well-developed trilaminar structure: the innermost layer — the tunica intima, the middle layer — the tunica media and the outer layer— adventitia [25, 30].

The early stage of atherosclerosis consists of the subendothelial accumulation of lipid-rich macrophages and T-lymphocytes within the innermost layer of the artery wall called the tunica intima, followed by proliferation and subendothelial migration of smooth muscle cells from the tunica media. This process leads to an increasing thickness of the tunica intima and media and in consequence to the development of more advanced complex lesions known as fibrous plaques [6, 19, 26, 27].

Until recently the early stages of atherosclerosis could only be observed in the artery wall during post mortem examinations. This is the reason for the emerging interest, among clinicians, in the possibility of monitoring the arterial wall structure in vivo.

This requirement for the assessment of early atherosclerotic changes has been met by high-resolution B-mode ultrasonography imaging of the carotid artery wall. In 1982 Zweibel [35] was the first to describe the ultrasonic image of the pattern of lines that constitute the intima and media complex of the large artery wall. A few years later Pignoli et al. described the “double line” pattern of the normal carotid artery wall corresponding to two echogenic interfaces: the first echogenic interface on the far wall is the lumen-intima interface and the second arises from the media-adventitia interface [25].

Most of the ultrasound studies were performed on the far wall of the carotid artery because when the near artery wall is imaged, the ultrasound beam passes first through the highly echogenic adventitia and then through the less echogenic intima-media complex, and finally through the weakly echogenic blood in the vessel lumen. The examination of the adventitia-media interface could be questionable due to technical reasons i.e. a not very clear picture of the structures [30]. Therefore, in the case of most of the studies the IMT measurement was performed on the far wall. Nevertheless, the improvement of ultrasound equipment with the improving image resolution took place and some new studies using cumulative far and near wall IMT measurement gave good results [16, 22].

This ultrasound technique with the use of high-resolution B-mode ultrasound is now one of the most important methods that provide information on the artery wall [1, 3, 11, 15, 32]. The early atherosclerotic changes within the subendothelial region of the artery wall are not accessible for in vivo examination with the use of other imaging methods such as MRI or angiography. The ultrasound method is non-invasive, safe and relatively inexpensive. The images obtained by this method can accurately be described as “ultrasonic biopsy” [5].

In order to apply ultrasound in clinical practice it is essential to demonstrate that the images observed on ultrasound examination reflect the real morphological structure. Therefore, substantial effort was applied in order to prove the concurrence between ultrasound and histological measurements.

On the basis of B-mode measurement of the distance between intimal and medial thickness, and correlation studies with the thickness of different combinations of tissue species of carotid arteries evaluated by gross and microscopic examination, Pignoli et al. concluded that the results of ultrasound examination of intimal and medial thickness did not differ significantly from the intimal plus medial thickness measured on histopathological examination [25]. This report was later validated in vitro [24, 34].

Recently, Schulte-Altedorneburg et al. [28] demonstrated the accuracy of in vivo carotid B-mode ultrasound compared with post mortem histopathological analysis in the same individuals. The study showed that B-mode ultrasound provides a reliable approach for the in vivo measurement of cross-sectional area. Measurements of IMT in vivo were systematically larger but the difference was proportional and maintained. The results indicated that the agreement between the two methods was good with a mean difference between histological and in vivo ultrasound examination of about 30%. Histological preparation of the specimen, the fixation in formalin, dehydration in ethanol, and embedding in paraffin may contribute to tissue shrinkage of even up to 30–40% [2].

Currently, it is accepted that IMT is a reliable and proportional representation of the real morphologic structure of the artery wall [28, 30].

Many studies show a strong correlation between IMT and atherosclerosis risk factors. IMT is associated with most of the known atherosclerotic risk factors, namely: arterial hypertension, dyslipidaemias, smoking, diabetes, increased body mass index, increased plasma concentration of homocysteine and the others [3, 11, 12, 18, 21]. IMT was also shown to have a linear relation to age and to the total number of risk factors [3, 32].

In a large prospective study O’Leary et al. [23] demonstrated that an increase in IMT of 1 standard deviation (in the study it was only 0.2 mm) in the
common and internal carotid arteries was associated with a 33 to 43% increase in the risk of stroke and with similar increases in the risk of myocardial infarction. After adjustment for age, sex and other risk factors assessed in the study, the risk rates were still significantly increased and were 24% and 28%, respectively [22].

Taking the above into consideration the measurement of the IMT of the carotid artery increased in importance and acceptance as an end-point in large clinical trials [8, 9, 13, 16, 20, 22, 23, 29]. Studies such as the Pravastatin, Lipids and Atherosclerosis Prevention Study (PLAC-II) [9], the Asymptomatic Carotid Artery Progression Study (ACAPS) [13], LIPID [20], and others have shown that long-term studies can be carried out successfully while maintaining a high reproducibility factor. It is also worth noting that the results of IMT measurement influence clinical decision-making in particular with regard to the initiation of more aggressive therapy (for example with statins) and when serial examinations are performed to assess the results of treatment. Thus, the results of carotid morphology measurements performed by different sonographers or retrieved from images stored — on discs or video-tapes — by various readers, must be reliable with a high reproducibility factor.

The aim of this study was to investigate the variability of ultrasonography in assessing the IMT of the carotid artery. The objective was to evaluate the reproducibility of the measurements made by the same investigator on two separate occasions (intraobserver variability) and the reproducibility of measurements between all four sonographers in the laboratory (interobserver variability). A second objective was to determine any influence of experience in ultrasound examination on inter- and intraobserver variability.

**MATERIAL AND METHODS**

IMT measurements were performed on 25 subjects (15 post stroke and 10 without perceptible vascular diseases) with a mean age 65 (SD ± 9) years, 12 of whom were female. Examination was performed with patients lying in the supine position. The subjects were examined using a Sonoline Sienna Duplex Scanner (Siemens) equipped with a 7.5 MHz linear transducer. The Colour Duplex mode was used for identification of the internal and external carotid artery. For the purpose of the evaluation, the carotid artery was divided into three segments based on arterial anatomy, similarly to previous studies [31] with some modifications. The first segment included the distal 2 cm of the vascular wall of the common carotid artery (CCA) immediately proximal to the dilatation of the bifurcation. The second segment, the carotid bulbs (BULB), was defined as the segment between the carotid dilatation and the flow divider of the internal and external carotid arteries. The third segment, the internal carotid artery (ICA), was defined as the segment immediately distal to the tip of the flow divider over a 1 cm distance. The loss of the parallel wall configuration, which marks the origin of the bulb segments, was easily identified as a consistent marker of the distal end of the CCA segment. The left and right arteries were scanned on longitudinal ultrasound image on the far walls of the arteries in artero-posterior and lateral planes. The IMT was defined as the distance between the leading edge of the first bright line of the far wall that consists of the lumen-intima interface and the leading edge of the second bright line that is the media-adventitia interface.

IMT was measured in end diastole, which was determined by the use of the 64-frame cinema system of the machine. When plaque was imaged the thickness of the plaque was taken as the IMT. The measurements were performed at a distance of 2 cm from the CCA, at 5 mm intervals (5 measurements in each plane), in the bulb — 4 measurements in equal intervals, and within 1 cm of the ICA — 3 measurements at 5 mm intervals in each plane. Additionally the maximum IMT in each region examined was assessed. In total 32 measurements were performed in each artery. All measurements were performed with the use of the machine’s electronic caliper with a resolution of 0.1 mm. All pictures were frozen and stored on 3.5-inch diskettes with the use of an electronic system attached to the machine. The measurements were repeated within a 3-week interval by sonographer “A”. Additionally three other readers with different levels of experience and skill in ultrasonography in the measurements were asked to perform the same measurements in duplicate with at least a 3-week interval between.

Sonographer “A” had 6 years’ experience in vascular ultrasound, “B” had 4 years’ experience, “C” had 3 years’, and reader “D” was a post-doctoral trainee, with 1 year of experience.

**Statistical analysis**

The accuracy of various readers was analysed using the Bland-Altman method [7]. The intraobserver variability was assessed for data from the first and second measurements of the IMT for each of the four readers, A, B, C and D. For the evaluation of the interobserver variability, measurements of readers...
B, C, and D were compared with measurements taken by reader A. The interobserver variability was assessed for both the first and the repeated IMT measurements. Bland-Altman analysis was performed both for the differences between measurements and the ratios of the differences divided by the average values of measurements. The results are presented as the bias between the two measurements with SD and 95% confidence intervals. Moreover, all measured values were compared using the non-parametric Wilcoxon test for paired data both for intra- and interobserver measurements. In addition, the association between the results of repeated IMT measurements for all readers was analysed using the non-parametric Spearman correlation. The results are presented as correlation r coefficient and p value.

RESULTS

Depending on the reader, the mean values of all measurements were between 0.93 and 1.04 mm, (Table 1).

Intraobserver variability

Intraobserver variability was in the range from 1 to 1.7%, and the maximal differences were 0.019 mm. This difference did not reach statistical significance (Table 2).

Interobserver variability

The relative difference of the interobserver variability for the first reading session was in the range 1.6 to 8.8%, whereas for the second reading session it was from 0.9 to 10.2%. The maximal values of absolute differences were in the range from 0.036 to 0.082 mm for the first reading session and from 0.012 to 0.08 for the second one. The interobserver variability was not significantly different only for comparison of reader A with reader B for both reading sessions (Table 4). The correlation coefficients for the interobserver variability ranged from 0.72 to 0.83 for the first reading session and from 0.72 to 0.82 for the second one (p < 0.0001 for all correlations) (Table 5).

The best (panel A) and the worst (panel B) interobserver variabilities are shown in Figure 1.

DISCUSSION

The study showed that the obtained inter- and intraobserver differences evaluated during repeated measurements of IMT were small and therefore the reproducibility of results is reliable.

Such variability could not have any influence on clinical decision-making. The interobserver variability for all pairs of investigators also shows concurrence and therefore does not have any impact on clinical decisions either. Even the greatest observed difference of 0.08 mm (Table 4) was within the limits of the method resolution i.e. 0.1 mm. The greatest difference of 0.11 mm in mean values of IMT for all measurements between readers “A” and “C” in the second reading session (Table 1) is also acceptable.

| Table 1. Mean values of IMT measured by all readers during both reading sessions |
|------------------------------------------|------------------------------------------|
|                                         | 1st measurement                               | 2nd measurement                               |
|                                         | Mean [mm] | SD [mm] | Mean [mm] | SD [mm] |
| Reader A                                 | 0.93      | 0.46    | 0.93      | 0.47    |
| Reader B                                 | 0.93      | 0.42    | 0.91      | 0.41    |
| Reader C                                 | 1.00      | 0.46    | 1.04      | 0.46    |
| Reader D                                 | 0.98      | 0.41    | 0.99      | 0.43    |

SD — standard deviation

There was a high correlation for each of the investigators between the two separate measurement sessions with r value ranging from 0.92 up to 0.95, p < 0.0001 for all the correlations (Table 3).

The best (panel A) and the worst (panel B) intraobserver variabilities are shown in Figure 1.

| Table 2. The intraobserver variability between two reading sessions |
|---------------------------------------------------------------|---------------------------------------------------------------|
|                                                             | Difference [mm] | Relative difference (%) |
|                                                             | Bias | SD | 95% CI       | Bias | SD | 95% CI |
| Reader A                                                    | 0.003 | 0.086 | −0.16 to 0.17 | 1.0 | 10.2 | −19.0 to −21.1 | NS |
| Reader B                                                    | 0.019 | 0.170 | −0.31 to 0.35 | 1.7 | 15.0 | −27.8 to −31.1 | NS |
| Reader C                                                    | 0.012 | 0.078 | −0.17 to 0.14 | −1.1 | 8.6 | −17.9 to −15.7 | NS |
| Reader D                                                    | 0.001 | 0.081 | −0.16 to 0.16 | −0.25 | 9.0 | −17.9 to −17.4 | NS |

SD — standard deviation, CI — confidence interval
Table 3. The intraobserver correlation for each reader

<table>
<thead>
<tr>
<th>Reader</th>
<th>Correlation coefficient</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1—A2</td>
<td>0.95</td>
<td>p &lt; 0.0001</td>
</tr>
<tr>
<td>B1—B2</td>
<td>0.92</td>
<td>p &lt; 0.0001</td>
</tr>
<tr>
<td>C1—C2</td>
<td>0.95</td>
<td>p &lt; 0.0001</td>
</tr>
<tr>
<td>D1—D2</td>
<td>0.95</td>
<td>p &lt; 0.0001</td>
</tr>
</tbody>
</table>

Table 5. The interobserver correlation between reader A and readers B, C and D for both reading sessions

<table>
<thead>
<tr>
<th>Reader</th>
<th>1st reading</th>
<th>2nd reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reader B</td>
<td>0.75; p &lt; 0.0001</td>
<td>0.72; p &lt; 0.0001</td>
</tr>
<tr>
<td>Reader C</td>
<td>0.72; p &lt; 0.0001</td>
<td>0.83; p &lt; 0.0001</td>
</tr>
<tr>
<td>Reader D</td>
<td>0.80; p &lt; 0.0001</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

Table 4. The interobserver variability of readers B, C, and D compared with reader A for both reading sessions

<table>
<thead>
<tr>
<th>Readers A vs.</th>
<th>Difference [mm]</th>
<th>95% CI</th>
<th>Relative difference (%)</th>
<th>Bias</th>
<th>SD</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 vs. B1</td>
<td>−0.036</td>
<td>−0.53 – 0.52</td>
<td>−1.6</td>
<td>25.7</td>
<td>−51.9 – 48.8</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>A1 vs. C1</td>
<td>−0.043</td>
<td>−0.55 – 0.46</td>
<td>−6.7</td>
<td>24.5</td>
<td>−54.7 – 41.3</td>
<td>0.0026</td>
<td></td>
</tr>
<tr>
<td>A1 vs. D1</td>
<td>−0.082</td>
<td>−0.53 – 0.37</td>
<td>−8.8</td>
<td>19.9</td>
<td>−47.9 – 30.2</td>
<td>&lt; 0.0001</td>
<td></td>
</tr>
<tr>
<td>A2 vs. B2</td>
<td>−0.012</td>
<td>−0.42 – 0.44</td>
<td>−0.9</td>
<td>23.0</td>
<td>−46.0 – 44.2</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>A2 vs. C2</td>
<td>0.058</td>
<td>−0.56 – 0.45</td>
<td>−8.5</td>
<td>25.1</td>
<td>−57.7 – 40.6</td>
<td>0.0017</td>
<td></td>
</tr>
<tr>
<td>A2 vs. D2</td>
<td>−0.080</td>
<td>−0.55 – 0.38</td>
<td>−10.2</td>
<td>22.1</td>
<td>−53.4 – 33.1</td>
<td>&lt; 0.0001</td>
<td></td>
</tr>
</tbody>
</table>

SD — standard deviation, CI — confidence interval
It is worth noting that the smallest difference detected was for the two most experienced investigators (“A” and “B”). The interobserver variability increased with the decrease in investigator experience, which confirms that the reproducibility and variability of carotid morphological measurements increase with experience of the investigator.

The limits of concurrence for inter- and intraobserver comparison, shown in Figures 1 and 2, are relatively small and acceptable from the clinical point of view. Even after a training period of one year an investigator using the ultrasound method was able to obtain relevant results that did not influence significantly clinical assessments or decisions. The obvious limitation of our study lies in the fact that the morphological measurements were obtained from frozen images stored on diskettes. Moreover, the majority of clinical trials and multicentre investigations opt for the same procedure. In most of them the images were stored on video tape and sent to the reading centres, where readers blind to the clinical data performed the IMT measurements off-line [8, 22, 29]. On the other hand most of the in vivo studies compared the measurements of two investigators, including four investigators, each patient had to be examined eight times in a relatively short time. Performing such IMT measurements on the same patient makes the study impossible. Our experience indicates that the most troublesome is the appropriate identification of the lines that constitute the intima-media complex and also the precision of measurement.

In the advent of electronic storage of the morphological images, the situation that two or more investigators will assess the same images will become more common.

Our results are similar to those achieved by the other authors. In the literature the mean differences of IMT measurements vary from even 0.04 mm to 0.66 mm. The majority of studies achieve a mean difference in the range of 0.04 mm to 0.3 mm. For the intraobserver variability the differences were less and most of them varied between 0.03 mm and 0.1 mm or even reached 0.007 mm when an automated method was used [4, 17, 33]. It is worth noting that despite the use of manual measurement the difference in the best measurements made by the most experienced pair of investigators was similar to that achieved with the use of the digital automated system (from 0.007 mm to 0.012 mm in automated measurements [4, 33] v. 0.012 mm in our study).

The future development of morphological measurement appears to be based on digitalisation of the systems. Currently more and more laboratories have at their disposal ultrasound machines with automatic digital systems. In an investigation of inter- and intraobserver reproducibility, it was demonstrated that automated IMT measurements can reduce the reproducibility error even by more than 50%, and obtain a shorter measurement time [10, 14, 30, 33].

**CONCLUSIONS**

The analogue measurements performed by our team show reliable reproducibility in terms of the results of morphologic measurements. The differences obtained in the study were less than the error of the method and therefore should not influence clinical decision-making.

Additionally, this study demonstrated that interobserver concurrence increases with the increasing experience of the investigators.

**REFERENCES**


