Ultrasound semiautomatic versus manual estimation of carotid intima-media thickness: reproducibility and cardiovascular risk stratification

Caterina Beatrice Monti¹, Marco Ali², Davide Capra¹, Federico Wiedenmann³, Giulia Lastella⁴, Francesco Secchi¹,⁵, Francesco Sardanelli¹,⁵

¹Department of Biomedical Sciences for Health, Università degli Studi di Milano, ²Unit of Diagnostic Imaging and Stereotactic Radiosurgery, Centro Diagnostico Italiano, ³Medicine and Surgery School, Università degli Studi di Milano, ⁴Postgraduation School in Radiodiagnostics, Università degli Studi di Milano, ⁵Unit of Radiology, IRCCS Policlinico San Donato, Milan, Italy

Abstract

Aims: Carotid intima-media thickness (CIMT) is used increasingly as an imaging biomarker of cardiovascular risk (CVR). Our aim was to compare semiautomatic CIMT (sCIMT) versus manual CIMT (mCIMT) for reproducibility and prediction of CVR. Materials and methods: Two independent readers measured sCIMT and mCIMT on previously acquired images of the right common carotid artery of 200 consecutive patients. Measurements were performed twice, four weeks apart; sCIMT was reported along with an image quality index (IQI) provided by the software. CVR stratification was compared for thresholds established by mCIMT studies, adapted for sCIMT according to a regression model. Results: sCIMT (median 0.67 mm, interquartile range [IQR] 0.57–0.76 mm) was significantly lower (p<0.001) than mCIMT (median 0.76 mm, IQR 0.63–0.84 mm; p=0.832, ρ=0.832, p<0.001, slope 0.714, intercept 0.124). Overall, intra-reader reproducibility was 76% for sCIMT and 83% for mCIMT (p=0.002), inter-reader reproducibility 75% and 76%, respectively (p=0.316). In 129 cases with IQI≥0.65, reproducibility was significantly higher (p≤0.004) for sCIMT than for mCIMT (intra-reader 85% versus 83%, inter-reader 80% versus 77%). The agreement between sCIMT and mCIMT for CVR stratification was fair both overall (κ=0.270) and for IQI≥0.65 (κ=0.345), crude concordance being 79% and 88%, respectively. Conclusions: Reproducibility of sCIMT was not higher than mCIMT overall but sCIMT was significantly more reproducible than mCIMT for high-IQI cases. sCIMT cannot be used for CVR stratification due to fair concordance with mCIMT, even for high IQI. More research is required to improve image quality and define sCIMT-based thresholds for stratification of CVR.

Keywords: carotid intima-media thickness; Doppler; cardiovascular diseases; reproducibility of results; computer-assisted

Introduction

Stroke is a major cause of cardiovascular death, representing the second worldwide leading killer after ischemic heart disease [1]. In Europe, stroke has an age-standardized incidence of 95 to 290/100,000 per year, with a mortality ranging from 15% to 35% [2]. Up to now, different scores to assess the risk of cardiovascular events have been developed [3,4], but they are still not completely satisfactory: many clinical events still occur in patients considered to be at low or intermediate risk [5,6].

Carotid atherosclerosis is one of the main risk factors for ischemic stroke [7]. Carotid intima-media thickness (CIMT) represents the width of the innermost portion of the carotid artery wall and may increase as atherosclerosis progresses, due to deposition of lipids and intimal proliferation [8]. Thus, CIMT has emerged as a predictor of cardiovascular events: its measurement represents a class Ila recommendation for cardiovascular risk (CVR)
assessment, where CVR is defined as a risk for a healthy adult to develop cardiovascular events associated with atherosclerotic vascular disease [9]. The study of Polak et al [10], conducted on the Framingham cohort, found that a significantly higher number of patients with larger CIMT values reported cardiovascular disease at follow-up when compared to those with normal CIMT values. Nevertheless, the use of CIMT for cardiovascular risk assessment is still debated, since there are no standardized methods for its measurement and no established reference values for normality [11]. The most important technical discrepancies between reported methods of CIMT measurement may be found in the choice of the carotid segment to evaluate and in measurements which are reported, which could be means or maximums or any combination of the two [12].

In clinical practice, CIMT may be measured with a routine ultrasonography of the supra-aortic vessels; therefore, its assessment is easy, low-cost and non-invasive [13]. Semiautomatic methods for CIMT calculation have become increasingly popular in the last years, since they allow quick measurements and a supposedly higher precision, possibly overcoming the limitations of non-standardized manual evaluation [14].

The purpose of our study was to compare semiautomatic CIMT versus manual CIMT for reproducibility and prediction of cardiovascular risk.

Materials and methods

The local Ethics Committee approved this study (Ethics Committee of San Raffaele Clinical Research Hospital; protocol code CardioRetro; approved on March 9th, 2017; amended on May 10th, 2018). “This study was partially supported by Ricerca Corrente founding from Italian Ministry of Health to IRCCS Policlinico San Donato. There was no overlap with subjects from previous research performed at our Institution. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. Due to the retrospective design of this study, specific informed consent was waived.

Ultrasound B-mode images of consecutive patients who underwent a color-Doppler ultrasound of the supra-aortic vessels at our institution were retrospectively re-evaluated for CIMT assessment. All patients who underwent color-Doppler ultrasound of the supra-aortic vessels were included in the study, regardless of sex, age, or cardiovascular risk factors. Exclusion criteria were repeated examinations on the same patient, significant carotid stenosis or previous treatments for carotid stenosis.

Image acquisition

Ultrasound examinations were conducted using a RS80A Prestige (Samsung Healthcare, Seoul, South Korea) equipment with a 7.5-MHz linear transducer with a field of view of 50 mm. Patients were examined while lying supine or sitting down; when that was not an option, with a slightly extended neck. One image per patient of the right common carotid that permitted a good visualization of CIMT was stored.

Image processing

Average CIMT of the right common carotid artery was assessed twice, 4 weeks apart, by a radiologist with ten years of experience in carotid ultrasound, and once by a second radiologist with two years of experience, with two different methods: semiautomatic and manual. The first set of measurements of the more experienced reader along with the other two sets of measurements were used for the assessment of reproducibility.

A dedicated semiautomatic software (AutoIMT+, Samsung Healthcare, Seoul, South Korea) was utilized for semiautomatic CIMT calculations. When such software failed to identify the common carotid artery, for instance measuring a section of muscle, readers could manually correct it by selecting the center of the desired vessel. AutoIMT+ software provided CIMT measurement, the number of points used for CIMT calculation and a so-called “quality index” (QI), related to the ratio between length of the measured segment and number of points, which were both reported along with CIMT.

For manual CIMT calculation, each radiologist took three different measurements, as recommended by the Italian Society for Vascular Investigation [11,12,15] using a dedicated software for image viewing. Such measurements were then averaged.

A comparison between semiautomatic and manual measurements is presented in figure 1.

Statistical analysis

Continuous variables were tested for normality with the Shapiro-Wilk test. Parametric data were reported as

![Fig 1. Comparison between semiautomatic and manual carotid intima-media thickness measurements in a 73-year-old female.](image-url)
a mean ± standard deviation (SD), while non-parametric
data were reported as a median and interquartile range.

Intra- and inter-reader reproducibility of both semi-
amtomatic and manual CIMT measurements were as-
sessed with Bland-Altman analysis, and it was reported
as a complement to one of the ratio between two times
the SD and mean measure, along with the coefficient of
reproducibility (CoR) and bias. Differences between data
reproducibility were assessed with the F-test of equality
of variances.

Differences between semiautomatic and manual data
were assessed with the t-test when variables were nor-
mlly or near-normally distributed, or by the Wilcoxon
test when variables were not normally or near-normally
distributed. Correlations between semiautomatic and
manual data as well as between image quality and points
of semiautomatic measurements were appraised with
Pearson or Spearman correlation with regards to data
distribution. Correlation coefficients were interpreted
according to Evans [16]. Concordance between semi-
amtomatic and manual CIMT with regards to cardiovas-
cular risk stratification were assessed with Cohen κ as a
measure of agreement, as well as reporting the percent-
age of crude concordant cases. Thresholds for manual
CIMT were chosen according to Lorenz et al [17] and
adapted according to a regression model: namely the cor-
relation between semiautomatic and manual CIMT was
modelized, and thresholds for semiautomatic CIMT were
obtained from the relation between the two. CIMT was
considered positive for cardiovascular risk when equal to
or above threshold, negative when below.

Subgroup analyses were performed according to the
value of QI, using different QI thresholds and calculating
both manual and semiautomatic CIMT reproducibility by
only including patients whose semiautomatic QIs were
equal to or above the desired thresholds. If overcoming
one particular QI threshold, semiautomatic CIMT repro-
ducibility went from lower to higher than that of manual
CIMT, such threshold was used for further subgroup
analysis.

Statistical analysis was performed with SPSS v22
(IBM SPSS Inc., Chicago, IL, USA), and MATLAB
R2018b (Mathworks, Natick, MA, USA); P-values <0.05
were considered statistically significant.

Results

Overall results
The color Doppler examinations of supra-aortic ves-
sels of 200 consecutive patients (median 71 years, inter-
quartile range [IQR] 63–78 years), conducted from May
2017 to December 2017 at our institution, were retro-
spectively analyzed. Among these patients, 103 (51%)
were females. The characteristics of the study population
are reported in Table I.

The software AutoIMT+ failed to properly detect the
common carotid artery on the stored image in 15% of
patients (N=30). In such cases, manual adjustments were
made: another point at the center of the vessel was cho-

Table I. Study population characteristics

<table>
<thead>
<tr>
<th>Patients characteristics (n=200)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
</tr>
<tr>
<td>Females (n,%)</td>
</tr>
<tr>
<td>Manual CIMT (mm)</td>
</tr>
<tr>
<td>Semiautomatic CIMT (mm)</td>
</tr>
</tbody>
</table>

Data are reported as median and interquartile range. CIMT: carotid
intima-media thickness

Fig 2. Correlation between semiautomatic and manual carotid
intima-media thickness (CIMT), ρ=0.832, p<0.001, modeled
into a regression line with a slope of 0.714 and an intercept of
0.124. This regression model was then used for translating
thresholds for manual CIMT to use them for semiautomatic
CIMT.
Semiautomatic and manual CIMT agreed on 144 negative cases and 13 positive cases, 35 cases were positive for semiautomatic CIMT and negative for manual CIMT, and 8 vice versa. Agreement for risk prediction was fair (κ=0.270), with a crude concordance of 79%.

Intra-reader reproducibility on the whole study sample was 76% with a bias of 0.003 mm and CoR of 0.163 mm for semiautomatic CIMT assessment and 83% with a bias of -0.014 mm and CoR of 0.130 mm for manual CIMT calculation. Inter-reader reproducibility on the whole study sample was 75% with a bias of -0.001 mm and CoR of 0.170 mm for semiautomatic CIMT assessment and 76% with a bias of 0.024 mm and CoR of 0.130 mm for manual CIMT calculation. Intra-reader reproducibility was significantly higher for manual CIMT (p=0.002), while inter-reader reproducibility was not significantly different (p=0.316) for semiautomatic CIMT than manual CIMT. Bland-Altman plots are shown in figure 3.

A strong positive correlation (ρ=0.678, p<0.001) was found between QI (0.86, 0.79–1.00) and points of length of the section used for measurements (115 points, 81–157 points), between QI and inter-reader reproducibility (ρ=0.724, p=0.005), and between QI and inter-reader reproducibility (ρ=0.900, p<0.001).

**Subgroup analysis**

Results of subgroup analyses for reproducibility according to QI were reported in Table II. The subgroup having a semiautomatic CIMT assessment with QI≥0.65 was composed by 129 subjects, 59 males (46%) and 70 females (54%), whose median age was 70 years (IQR 61–77 years). There were no significant differences in terms of sex (p=0.652) or age (p=0.882) when compared to the entire population. At QI≥0.65, intra-reader reproducibility was 85% with a bias of -0.004 mm and CoR of 0.098 mm for semiautomatic CIMT and 83% with a bias of -0.012 mm and CoR 0.126 mm for manual CIMT. Inter-reader reproducibility was 80% with a bias of 0.003 mm and CoR of 0.013 mm for semiautomatic CIMT and 77% with a bias of 0.015 mm and CoR of 0.172 mm for manual CIMT. Both intra- and inter-reader reproducibilities are on the left; inter-reader reproducibilities are on the right.

**Fig 3.** Bland-Altman charts representing the reproducibility of (from top line to the bottom line) manual carotid intima-media thickness (CIMT) over the whole population, overall semiautomatic CIMT, manual CIMT with quality index ≥0.65, and semiautomatic CIMT with quality index ≥0.65. Intra-reader reproducibilities are on the left; inter-reader on the right.

**Fig 4.** Reproducibility values according to different quality index thresholds. The yellow line represents manual intra-reader reproducibility, the blue line manual inter-reader reproducibility, the purple line semiautomatic inter-reader reproducibility, and the red line semiautomatic intra-reader reproducibility. At a QI threshold ≥0.65, it may be observed that both semiautomatic intra- and inter-reader reproducibility overtakes manual intra- and inter-reader reproducibility, respectively (the purple line crossing the blue line, and the red line crossing the yellow line). Moreover, it may be noticed that manual reproducibility values tend to remain constant with regards to QI, semiautomatic values encounter a steep growth as the QI rises.
reproducibility of semiautomatic CIMT were significantly higher (p=0.004 and p=0.006, respectively) than those of manual CIMT. The correlation between QI and points led to the correspondence of a threshold of 24 points for QI ≥0.65. A chart depicting the variations of reproducibility values according to QI thresholds may be seen in figure 4.

In this subgroup, while the CIMT semiautomatic values (0.66 mm, IQR 0.55–0.75 mm) were significantly lower (p<0.001) than manual CIMT values (0.75 mm, IQR 0.64–0.83 mm), a very strong positive correlation (p=0.867, p<0.001) persisted between semiautomatic and manual CIMT assessment. Semiautomatic CIMT detected 17 high-risk patients, while manual CIMT showed 8 patients at a high cardiovascular risk. Semiautomatic and manual CIMT agreed on 109 negative cases and 5 positive cases, while 12 cases were positive according to semiautomatic CIMT and negative according to manual CIMT, and 3 vice versa. Agreement for risk prediction between the two measurements was fair (κ=0.345), with a crude concordance of 88%.

### Discussion

The aim of our study was to compare semiautomatic CIMT versus manual CIMT for reproducibility and prediction of cardiovascular risk. Overall, semiautomatic CIMT resulted in being not more reproducible than by manual CIMT. However, in the cases with high image quality, reproducibility of semiautomatic CIMT was significantly higher for that of manual CIMT. In addition, the agreement between semiautomatic CIMT and manual CIMT for cardiovascular risk (CVR) stratification was fair both overall and for cases with a high image quality.

Our results should be extendable to the whole Caucasian target population. In fact, the median age of our entire population reflects well that of those at a high risk of stroke [2] and manual CIMT values from our population are compatible with their age [18].

We showed a high precision of manual CIMT: its overall reproducibility is highly satisfactory (intra-reader 83%, inter-reader 81%). Thus, we confirm the results by Plasencia Martinez JM, et al. [19] who found a good

<table>
<thead>
<tr>
<th>QI threshold</th>
<th>№ of patients</th>
<th>Reproducibility (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Intra-observer</td>
<td>Inter-observer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>manual</td>
<td>manual</td>
</tr>
<tr>
<td>0.00</td>
<td>200</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>0.05</td>
<td>199</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>0.10</td>
<td>198</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>0.15</td>
<td>196</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>0.20</td>
<td>193</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>0.25</td>
<td>188</td>
<td>82</td>
<td>76</td>
</tr>
<tr>
<td>0.30</td>
<td>185</td>
<td>82</td>
<td>76</td>
</tr>
<tr>
<td>0.35</td>
<td>181</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>0.40</td>
<td>174</td>
<td>82</td>
<td>76</td>
</tr>
<tr>
<td>0.45</td>
<td>164</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>0.50</td>
<td>156</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>0.55</td>
<td>148</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>0.60</td>
<td>141</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>0.65</td>
<td>129</td>
<td>83</td>
<td>77</td>
</tr>
<tr>
<td>0.70</td>
<td>122</td>
<td>83</td>
<td>77</td>
</tr>
<tr>
<td>0.75</td>
<td>110</td>
<td>83</td>
<td>77</td>
</tr>
<tr>
<td>0.80</td>
<td>104</td>
<td>83</td>
<td>78</td>
</tr>
<tr>
<td>0.85</td>
<td>95</td>
<td>83</td>
<td>78</td>
</tr>
<tr>
<td>0.90</td>
<td>90</td>
<td>83</td>
<td>78</td>
</tr>
<tr>
<td>0.95</td>
<td>79</td>
<td>83</td>
<td>77</td>
</tr>
<tr>
<td>1.00</td>
<td>64</td>
<td>83</td>
<td>76</td>
</tr>
</tbody>
</table>

Table II. Reproducibility of measurements of the right carotid intima-media thickness (CIMT) for different quality index (QI) thresholds. Bold data indicates significantly lower reproducibility for semiautomatic CIMT, while italic data indicates significantly higher reproducibility for semiautomatic CIMT. Note that both intra- and inter-reader reproducibility become significantly higher for semiautomatic CIMT than manual CIMT when the QI threshold is ≥0.65.
inter-reader reproducibility (intraclass correlation coefficient 79%), and a good inter-reader reproducibility (intraclass correlation coefficient 72%). In addition, on our whole population, manual CIMT calculation seemed to be more precise than semiautomatic CIMT.

The significant positive correlation between semiautomatic and manual CIMT measurements on our entire population hints that both values are related to the actual value of CIMT. However, the agreement between the two measurements in attributing CVR according to currently used thresholds, derived from manual measurements [17], is only fair. Even with a crude concordance of 79% between semiautomatic and manual CIMT measurements in the attribution of CVR, agreement according to Cohen κ was low (κ=0.270), due to an unbalanced distribution of positive and negative cases (the higher number of negative cases compared to positive cases has a high impact on the calculation of κ). This might possibly be caused by the fact that manual CIMT calculation relies on three different measurements taken at different points, while automatic or semiautomatic CIMT measurements use a higher number of points on a whole section of the carotid artery. Therefore, manual averaging may overestimate the length of carotid portions with increased CIMT, since more remarkable points with a higher CIMT tend to be chosen by readers, and thus it may provide a less reliable estimate. Given that the thresholds for CVR derived from manual measurements, semiautomatic measurements may yield a low agreement with manual measurements for risk prediction.

Our positive correlation between QI and length of the analyzed carotid segment can be explained taking into account that an average measure calculated on a greater number of data provides a more reliable estimate. The significant positive correlation between reproducibility and QI suggested that CIMT values obtained on images with a higher QI could be more precise and reproducible. In fact, at a subgroup analysis of cases with a QI≥0.65, neither age nor sex had different distributions compared to the entire population, meaning that the higher reproducibility in the chosen subgroup was likely not attributable to differences in such characteristics. Both intra- and inter-reader reproducibility of the semiautomatic method on the subgroup with QI≥0.65 were increased to the point of being significantly higher than those of the manual method on the same subgroup. Therefore, semiautomatic CIMT quantification could be more precise than manual estimates, provided that more than 24 points are utilized.

Despite the increase in reproducibility observed using the semiautomatic method taking QI into account, compared to manual CIMT, agreement for risk prediction between semiautomatic and manual CIMT was still low also in the subgroup with QI≥0.65 (κ = 0.345), thus these two methods do not appear interchangeable for CVR prediction, also in the presence of high QI. However, we must consider that there are no reference thresholds for semiautomatic CIMT for CVR assessment, the only available thresholds being derived from manual CIMT measurements [17,18].

This study has limitations, the first being its retrospective monocentric design which means that results are only extendable to a population similar to the one referring to our center. Second, we used a series of images already acquired and stored in our database. This means that an important source of variability (that due to the repetition of the examination) has been taken out of the study design to avoid a confounding factor but this implies that we do not know what happens when the examination is repeated using manual or semiautomatic CIMT measurements. Third, due to the advanced age of some patients, at times the range of motion of their neck was not ideal and did not allow for the best possible CIMT visualization. However, the target age for the use of CIMT as a risk assessment is comparable to that of our population, therefore we expect this limitation to be present in all studies and in clinical practice. Another limitation may be represented by the fact that we appraised CIMT on the right common carotid artery due to it being common practice and for the benefit of continuity. Nevertheless, this approach was already suggested by a previous work by Rosvall et al [20], who also noted that a study by Lorentz et al [21] reported comparable prognostic significance.

In conclusion, using a series of carotid ultrasound images already acquired in a consecutive series of patients, we showed that CIMT measurement using semiautomatic method (AutoIMT+) has a significantly higher reproducibility than values obtained with manual methods only when considering images with QI≥0.65. In addition, due to low concordance, semiautomatic CIMT assessment with AutoIMT+ cannot substitute for manual calculations in the attribution of cardiovascular risk using a threshold derived from studies which used manual CIMT estimate. To exploit the role of semiautomatic CIMT, prospective studies should be conducted assessing the predictive value of semiautomatic CIMT measurement, i.e. using thresholds obtained with semiautomatic measurement.

**Conflict of interest:** none

**References**