Advances in imaging and endoscopic therapy in Barrett’s esophagus
Alvarez Herrero, L.

Citation for published version (APA):
Alvarez Herrero, L. (2014). Advances in imaging and endoscopic therapy in Barrett’s esophagus

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Validation of the Prague C&M classification in Barrett’s esophagus in clinical practice


Endoscopy. 2013 Nov;45(11):876-82
Abstract

**Introduction:** The Prague C&M classification to describe Barrett's esophagus (BE) length has found widespread acceptance but has only been validated by Barrett's experts scoring video sequences. Until now validation is lacking for its application in routine practice during real-time endoscopy. **Aim:** To evaluate agreement of BE and hiatal hernia (HH) length according to this classification by Barrett's experts and community hospital endoscopists, during real-time endoscopy. **Methods:** Patients underwent two consecutive endoscopies performed by different endoscopists. The study was performed in two cohorts: one cohort was seen by Barrett's experts the other cohort by community hospital endoscopists. Landmarks were recorded according to the Prague classification. Outcomes were inter-observer agreement (assessed with intra-class correlation coefficient [ICC]), absolute agreement and relative agreement. **Results:** 187 patients were included with median C3M5cm(IQR C1-7M4-9) BE and 3cm (IQR 2-5) HH length. ICC was 0.91(95%CI 0.88-0.93) for M, 0.92(0.90-0.94) for C and 0.59(0.49-0.68) for HH length. Absolute agreement ≤1cm was 74%(95%CI 68-80%) for C, 68%(62-75%) for M, and 63%(56-70%) for HH length. Relative agreement was 91% for C&M and 80% for HH length. Barrett's experts and community hospital endoscopists showed no differences in agreement. BE≤5cm had lower agreement compared to BE>5cm. **Conclusions:** Agreement was good for BE and reasonable for HH length. These findings strengthen the value of the Prague C&M classification to describe BE and HH length. Although absolute agreement during real-time endoscopy is high one should anticipate that C&M values length may vary by 1-2cm between two endoscopies.
Introduction

Barrett’s esophagus is a condition of the distal esophagus in which the normal squamous epithelium has been replaced by columnar lined epithelium containing intestinal metaplasia upon biopsy.1 This condition is a premalignant condition in which esophageal adenocarcinoma can arise through a multi-step transition sequence.2 Because of this risk, patients with Barrett’s esophagus are offered periodic surveillance endoscopies with biopsies.1, 3, 4

In addition to assessing the Barrett’s segment for neoplastic lesions, it is of clinical relevance to adequately document the Barrett’s length in a reproducible manner during surveillance endoscopies. Previous studies have suggested that longer Barrett’s segments are at higher risk of developing neoplasia.5-9 Assessment and registration of Barrett’s length might therefore be helpful in the future for cancer risk stratification, e.g. by lengthening or shortening the surveillance intervals according to the length of the Barrett’s segments. Furthermore, the Barrett’s length is also important when planning endoscopic therapy as well as for a more objective evaluation of the response to endoscopic therapy.

Recently the Prague C&M classification has been introduced as an easy and uniform way to describe the length of the Barrett’s segment during endoscopy.10 In this classification, the circumferential (C) and maximum (M) extent of the Barrett’s epithelium are described in centimeters. Since its introduction, the Prague C&M classification has found widespread clinical application. This classification has, however, only been validated through the scoring of video sequences by endoscopists with special interest in Barrett’s esophagus. To date, no data are available on its agreement during real-time endoscopy in routine practice.

The aim of this study was to evaluate in a post hoc analysis the agreement of Barrett’s esophagus and hiatal hernia length according to the Prague C&M classification by Barrett’s expert and community hospital endoscopists, during real-time endoscopy assessed at two consecutive endoscopic procedures.

Methods

Setting

We performed a post hoc analysis by analyzing the data from two previously published randomized cross-over trials comparing standard video endoscopy (SVE) with endoscopic tri-modal imaging (ETMI, an endoscopy system combining high-resolution endoscopy, autofluorescence endoscopy, and narrow band imaging) for detection of neoplasia in Barrett’s esophagus.11, 12 The design of
the studies required that all patients underwent two endoscopies on two separate occasions by different endoscopists. The first study enrolled 87 Barrett's patients who were referred for endoscopic work-up of endoscopically inconspicuous high grade intra-epithelial neoplasia (HGIN) or early cancer. The procedures in this study were performed at five international tertiary referral centers by eight endoscopists with expertise in imaging and therapy of Barrett's neoplasia. The second study included 100 patients with a confirmed histological diagnosis of low grade intra-epithelial neoplasia (LGIN). The procedures were performed at eight community hospitals in the Amsterdam region by nine endoscopists without specific interest in Barrett's esophagus imaging. Both trials were approved by the Institutional Review Boards of the participating centers (ISRCTN68328077 and ISRCTN91816824).

**Patient selection**

Inclusion criteria for both trials were: 1) age >18 years; 2) prior diagnosis of Barrett's esophagus defined as the presence of columnar lined epithelium in the tubular esophagus with specialized intestinal metaplasia on histological examination; 3) written informed consent. The study population in the tertiary referral centers consisted of patients with HGIN or early cancer (amenable for endoscopic therapy) in a Barrett's segment of at least C≥2M≥2 cm or C<2M≥4 cm in length (group I). The study population in the community hospitals consisted of patients with any Barrett's length with a prior diagnosis of LGIN confirmed by an expert gastro-intestinal pathologist (group II).

Exclusion criteria for both trials were: 1) presence of an advanced neoplastic lesion at the first endoscopy not allowing a delay in intervention for a period of 6 weeks (i.e. the minimum interval between the two cross-over endoscopies); 2) presence of active erosive esophagitis grade B or worse according to the Los Angeles classification of erosive esophagitis at the first endoscopy; 3) presence of conditions precluding histological sampling of the esophagus (e.g. esophageal varices, coagulation disorders, anticoagulant therapy).

**Study design and randomization**

All patients underwent two consecutive upper endoscopies. One procedure was performed with SVE (GIF-140, GIF-160 or GIFQ-160, Olympus, Hamburg, Germany) and the other procedure with ETMI (XGIF-Q240/260FZ, Olympus, Tokyo, Japan); all endoscopes have 5 cm interval markings on the shaft. The two endoscopic procedures were separated by an interval of 6-16 weeks and were performed by two different endoscopists who were blinded to the results of any previous endoscopies (i.e. study endoscopies as well as any previous endoscopy before the study). Endoscopists were assigned to the first procedure before randomization to SVE or ETMI.
Randomization of the technique was performed at the first endoscopy by opening a sealed opaque envelope. The second procedure was automatically performed using the alternative endoscopy technique by another endoscopist.

**Endoscopists**

All endoscopists received prior to the start of the study a training DVD containing instructions on how to use the Prague C&M classification. Endoscopists did not receive explicit instructions on the moment to read off the location of a landmark, while advancing or retrieving the endoscope. Also no explicit instructions were given on the deflation of the stomach for the measurement of the gastric folds location.

Patients in group I were seen by eight endoscopists with extensive expertise in endoscopic imaging and treatment of the Barrett’s esophagus (Barrett’s experts). These eight expert endoscopists formed four fixed pairs. Patients in group II were seen by nine different endoscopists without extensive expertise with Barrett’s esophagus (community hospital endoscopists). No fixed pairs of community hospital endoscopists were formed.

**Endoscopic procedure**

In all procedures, patients were sedated according to the standard protocol of the participating centers, mostly intravenous midazolam supplemented with fentanyl or pethidine if necessary. According to the Prague C&M classification, distance from the incisors to the diaphragmatic pinch, the top of the gastric folds and the circumferential and maximal extent of the Barrett’s segment were prospectively collected at both endoscopic procedures in each patient. In case of ETMI, these data were collected in the high resolution white light mode. The length of the Barrett’s segment was the distance between the top of the gastric folds and the most proximal circumferential Barrett’s extent (C) and between the top of the gastric folds and the most proximal (i.e. maximal) Barrett’s extent (M). Hiatal hernia length was the distance between the diaphragmatic pinch and the top of the gastric folds.

**Outcome parameters**

The primary outcome parameter was the agreement for C, M and hiatal hernia length, defined as:

a. inter-observer agreement for C, M and hiatal hernia length;

b. absolute agreement for C, M and hiatal hernia length;

c. and relative agreement for C, M and hiatal hernia length.
The secondary outcome parameters were the inter-observer agreement and absolute agreement for the different landmarks (diaphragmatic pinch, top of the gastric folds, C location and M location) measured as distance from the incisors in cm.

Subanalyses of data were performed to find differences in inter-observer agreement, absolute agreement and relative agreement for C and M length between shorter Barrett’s segments (mean M≤5 cm) and longer Barrett’s segments (mean M>5 cm); and differences in inter-observer agreement, absolute agreement and relative agreement for C, M and hiatal hernia length between Barrett’s experts and community hospital endoscopists.

**Statistical analysis**

Statistical analysis was performed with the Statistical Software Package version 16.0.2 for Windows (SPSS, Chicago, Illinois, USA). For descriptive statistics, means and standard deviations were used for variables with a normal distribution, and medians and interquartile ranges (IQR) were used for variables with a skewed distribution. Categorical data were compared using the Chi square test (the Fisher exact test when expected cell values were too small) or were tested for trend. Continuous data that was normally distributed were compared using the (paired) T test and ANOVA when multiple variables were tested. Continuous data that was not normally distributed were compared using the Mann-Whitney U test. Due to multiple testing Bonferroni correction was applied and a p-value of ≤0.01 was considered significant.

Inter-observer agreement was assessed with the intra-class correlation coefficient (ICC) considering an ICC value of 1.00 as perfect agreement and an ICC value of 0 as poor agreement. Comparison of the ICC between two groups was performed with the two-sample independent T-test on OpenEpi statistics (http://www.openepi.com/OE2.3/menu/openEpiMenu.htm).

ICC is mathematically defined as the ratio of the variance between the raters (i.e. the consecutive endoscopists) and the total variance of the study population. Thus with a given variance between endoscopists and a small variance in Barrett’s length between patients, the ICC will be low, while the same variance between endoscopists in a group of patients with a wider range of Barrett’s lengths will result in a higher ICC. Therefore we also report the absolute agreement and the relative agreement.

Absolute agreement for each variable (M, C or hiatal hernia length) was calculated for 0 cm discrepancy between both endoscopies, ≤1 cm discrepancy between both endoscopies, and ≤2 cm discrepancy between both endoscopies. Confidence intervals and differences with their confidence intervals for absolute agreement were calculated with the Confidence Interval Analysis package. As the ICC value may overestimate the agreement because of the wide range of C,M and hiatal hernia length within the patient population and absolute agreement
is not related to the length of the Barrett’s segment we created the relative agreement trying to overcome these drawbacks. Relative agreement was calculated for each variable (M, C or hiatal hernia length) by dividing the concordance between both endoscopies by the total length of the variable (mean of both endoscopies) for each patient. Concordance was calculated as: \(((\text{length on first endoscopy} + \text{length on second endoscopy}) - \text{difference between both endoscopies})/2\). For example if the first endoscopy resulted in \(M=6\text{cm}\), and the second in \(M=4\text{cm}\), the difference between both endoscopies is 2 cm and concordance was calculated as \((6\text{cm}+4\text{cm}-2\text{cm})/2=4\text{cm}\). Relative agreement in this case resulted in \(4\text{cm}(\text{concordance})/5\text{cm}(\text{mean of both endoscopies})\). In cases that both endoscopies showed a length of 0 cm for a variable, a relative agreement of 100% was set.

Finally, Bland-Altman plots were performed for each variable (M, C or hiatal hernia length) to visualize the relation between the mean length of each variable and its difference.$^{14}$

### Results

#### Baseline characteristics

Baseline characteristics are shown in Table 1. Compared to group I, group II consisted of slightly younger patients, with a shorter Barrett’s segment, and lower grades of neoplasia: all reflecting the selection criteria of the study.

Barrett’s expert endoscopists performed a mean of 22 (range 15-38) procedures and community hospital endoscopists performed a mean of 21 (range 11-37) procedures.
Table 1 Baseline characteristics

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Group I</th>
<th>Group II</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases, no.</td>
<td>187</td>
<td>87</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Age, years (±SD)</td>
<td>64 (±10)</td>
<td>67 (±9)</td>
<td>62 (±10)</td>
<td>0.001</td>
</tr>
<tr>
<td>Gender, % Male</td>
<td>81%</td>
<td>82%</td>
<td>80%</td>
<td>0.781</td>
</tr>
<tr>
<td>C, median of both endoscopies, cm (IQR)</td>
<td>C3 (1-7)</td>
<td>C4 (2-9)</td>
<td>C2 (1-5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>M, median of both endoscopies, cm (IQR)</td>
<td>M5 (4-9)</td>
<td>M8 (5-10)</td>
<td>M5 (3-7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HH, median of both endoscopies, cm (IQR)</td>
<td>3 (2-5)</td>
<td>3 (2-5)</td>
<td>3 (2-4)</td>
<td>0.112</td>
</tr>
<tr>
<td>Interval between endoscopies, weeks (IQR)</td>
<td>8 (7-10)</td>
<td>8 (6-10)</td>
<td>9 (7-10)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Overall pathology, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No neoplasia</td>
<td>10%</td>
<td>3%</td>
<td>16%</td>
<td></td>
</tr>
<tr>
<td>Indefinite</td>
<td>10%</td>
<td>1%</td>
<td>18%</td>
<td></td>
</tr>
<tr>
<td>LGIN</td>
<td>38%</td>
<td>32%</td>
<td>43%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HGIN</td>
<td>24%</td>
<td>36%</td>
<td>14%</td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>33%</td>
<td>28%</td>
<td>9%</td>
<td></td>
</tr>
</tbody>
</table>

Group I patients with HGIN or cancer undergoing two consecutive endoscopies by different endoscopists with extensive expertise with Barrett’s esophagus Group II patients with LGIN undergoing two consecutive endoscopies by different endoscopists without specific interest in Barrett’s esophagus SD standard deviation; IQR interquartile range; C circumferential extent Barrett’s esophagus; M maximum extent Barrett’s esophagus; HH hiatal hernia; LGIN low grade intraepithelial neoplasia; HGIN high grade intraepithelial neoplasia.

Barrett’s esophagus and hiatal hernia length

Overall inter-observer agreement and absolute agreement defined as a discrepancy of 0 cm, ≤1 cm or ≤2 cm between both endoscopies are shown in Table 2. Overall inter-observer agreement was almost perfect for circumferential and maximal Barrett’s extent, and moderate for hiatal hernia length. Overall absolute agreement, defined as a discrepancy of 0 cm, varied between 29-39%, increasing to 63-74% and 84-91% for a discrepancy of ≤1 cm and ≤2 cm, respectively (Table 2). Overall median relative agreement was 91% (IQR 75-100%) for C length, 91% (IQR 82-100%) for M length and 80% (IQR 67-100%) for hiatal hernia length. Bland-Altman plots for C, M and hiatal hernia length are shown in Figure 1.

No systemic differences in mean C, M or hiatal hernia length were seen when comparing the first and second endoscopic procedures (p=0.473, 0.550 and 0.754 respectively for C, M and HH) or when comparing the procedures performed with SVE and ETMI (p=0.049, 0.069 and 0.210 respectively for C, M and HH). Also no difference is mean C, M and hiatal hernia length were seen when comparing the 50 first procedures with the last 50 procedures (p=0.054, 0.057 and 0.343 respectively for C, M and HH). Finally no systemic differences were seen in group I between
the four pairs of endoscopists (p=0.378, 0.151 and 0.027 respectively for C, M and HH) and no systemic differences were seen in group II between the nine observers (p=0.907, 0.974 and 0.837 respectively for C, M and HH).

Figure 1 Bland-Altman plots for C, M and hiatal hernia length (in cm). The middle line in the graphs corresponds to the mean length, the upper and lower lines in the graphs correspond to the length +2 standard deviation and -2 standard deviation respectively.

Table 2 Inter-observer agreement and absolute agreement for circumferential and maximum Barrett’s length and hiatal hernia length in 187 patients undergoing two consecutive endoscopies by different endoscopists.

<table>
<thead>
<tr>
<th></th>
<th>Overall ICC (95% CI)</th>
<th>0 cm discrepancy % (95% CI)</th>
<th>≤1 cm discrepancy % (95% CI)</th>
<th>≤2 cm discrepancy % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.92 (0.90-0.94)</td>
<td>39% (32-46%)</td>
<td>74% (68-80%)</td>
<td>89% (85-94%)</td>
</tr>
<tr>
<td>M</td>
<td>0.91 (0.88-0.93)</td>
<td>31% (24-37%)</td>
<td>68% (62-75%)</td>
<td>91% (86-95%)</td>
</tr>
<tr>
<td>HH</td>
<td>0.59 (0.49-0.68)</td>
<td>29% (22-35%)</td>
<td>63% (56-70%)</td>
<td>84% (79-89%)</td>
</tr>
</tbody>
</table>

ICC intra-class correlation coefficient; CI confidence interval; M maximum extent Barrett’s esophagus; C circumferential extent Barrett’s esophagus; HH hiatal hernia.
Landmarks
Inter-observer agreement and absolute agreement for landmark location are shown in Table 3. Inter-observer agreement for M, C and top of the gastric folds location was almost perfect and for the diaphragmatic pinch location substantial. Absolute agreement defined as a discrepancy of 0 cm was 26-32%, 56-75% when defined as a discrepancy of ≤1 cm and 79-95% when defined as a discrepancy of ≤2 cm (Table 3).

Table 3 Inter-observer agreement and absolute agreement for landmark location in 187 patients with Barrett’s esophagus undergoing two consecutive endoscopies by different endoscopists.

<table>
<thead>
<tr>
<th>Location</th>
<th>Overall ICC (95% CI)</th>
<th>0 cm discrepancy % (95% CI)</th>
<th>≤1 cm discrepancy % (95% CI)</th>
<th>≤2 cm discrepancy % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C location</td>
<td>0.95 (0.94-0.96)</td>
<td>32% (25-38%)</td>
<td>66% (60-73%)</td>
<td>87% (82-92%)</td>
</tr>
<tr>
<td>M location</td>
<td>0.96 (0.95-0.97)</td>
<td>30% (23-37%)</td>
<td>75% (69-81%)</td>
<td>95% (90-97%)</td>
</tr>
<tr>
<td>Gastric folds</td>
<td>0.86 (0.82-0.89)</td>
<td>32% (25-38%)</td>
<td>70% (64-77%)</td>
<td>89% (85-94%)</td>
</tr>
<tr>
<td>Diaphragmatic pinch</td>
<td>0.68 (0.60-0.75)</td>
<td>26% (20-33%)</td>
<td>56% (49-63%)</td>
<td>79% (73-85%)</td>
</tr>
</tbody>
</table>

ICC intra-class correlation coefficient; CI confidence interval; M maximum extent Barrett’s esophagus; C circumferential extent Barrett’s esophagus.

Subanalyses for Barrett’s length and endoscopic expertise
For shorter Barrett’s segments (mean M≤5 cm) inter-observer agreement for C and M length was significantly lower compared to longer Barrett’s segments (mean M>5 cm) (Table 4). Relative agreement for M length was significantly lower in shorter Barrett’s compared to longer Barrett’s: 86% (IQR 80-100%) and 93% (IQR 88-100%) (p=0.002), respectively. On the contrary, when testing for trend, absolute agreement was higher in shorter Barrett’s segments (Table 5).

Between endoscopists with extensive expertise with Barrett’s esophagus (group I) and community hospital endoscopists without this expertise (group II) no differences were found in inter-observer agreement for Barrett’s segment and hiatal hernia length (Table 6). Absolute and relative agreement for Barrett’s segment and hiatal hernia length were also not significantly different.

Table 4 Inter-observer agreement of short and long Barrett’s esophagus for circumferential and maximum Barrett’s length.

<table>
<thead>
<tr>
<th></th>
<th>M ≤5 cm ICC (95% CI)</th>
<th>M &gt;5 cm ICC (95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.64 (0.51-0.75)</td>
<td>0.88 (0.83-0.92)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>M</td>
<td>0.44 (0.24-0.57)</td>
<td>0.83 (0.76-0.89)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

ICC intra-class correlation coefficient; CI confidence interval; M maximum extent Barrett’s esophagus; C circumferential extent Barrett’s esophagus.
Table 5 Absolute agreement of short and long Barrett's esophagus for circumferential and maximum Barrett's length.

<table>
<thead>
<tr>
<th>discrepancy</th>
<th>M ≤5 cm %</th>
<th>M &gt;5 cm %</th>
<th>Difference % (95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 cm</td>
<td>49%</td>
<td>29%</td>
<td>20% (6.2-35%)</td>
<td></td>
</tr>
<tr>
<td>≤1 cm</td>
<td>84%</td>
<td>63%</td>
<td>21% (8.3-32.9%)</td>
<td>0.007</td>
</tr>
<tr>
<td>≤2 cm</td>
<td>96%</td>
<td>83%</td>
<td>13% (4.3-22%)</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 cm</td>
<td>35%</td>
<td>26%</td>
<td>9.3% (-3.8-22%)</td>
<td>0.041</td>
</tr>
<tr>
<td>≤1 cm</td>
<td>73%</td>
<td>63%</td>
<td>10% (-3.3-23%)</td>
<td></td>
</tr>
<tr>
<td>≤2 cm</td>
<td>97%</td>
<td>85%</td>
<td>12% (3.8-20%)</td>
<td></td>
</tr>
</tbody>
</table>

M maximum extent Barrett's esophagus; C circumferential extent Barrett's esophagus; HH hiatal hernia; CI confidence interval. Percentages were tested for trend.

Table 6 Inter-observer agreement subdivided for expertise of endoscopists for circumferential and maximum Barrett's length and hiatal hernia length.

<table>
<thead>
<tr>
<th>Group</th>
<th>ICC (95% CI)</th>
<th>Group</th>
<th>ICC (95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.91 (0.87-0.94)</td>
<td>C</td>
<td>0.92 (0.89-0.95)</td>
<td>0.278</td>
</tr>
<tr>
<td>M</td>
<td>0.91 (0.87-0.94)</td>
<td>M</td>
<td>0.90 (0.85-0.92)</td>
<td>0.509</td>
</tr>
<tr>
<td>HH</td>
<td>0.60 (0.44-0.72)</td>
<td>HH</td>
<td>0.59 (0.44-0.70)</td>
<td>0.970</td>
</tr>
</tbody>
</table>

Group I patients with HGIN or cancer undergoing two consecutive endoscopies by different endoscopists with extensive expertise with Barrett's esophagus Group II patients with LGIN undergoing two consecutive endoscopies by different endoscopists without specific interest in Barrett's esophagus ICC intra-class correlation coefficient; CI confidence interval; M maximum extent Barrett's esophagus; C circumferential extent Barrett's esophagus; HH hiatal hernia.

Discussion

In this study, the use of the Prague C&M classification during real-time endoscopy was associated with good agreement (i.e. inter-observer agreement, absolute agreement and relative agreement) for C&M Barrett's length and a reasonable agreement for hiatal hernia length. No differences in agreement between Barrett's experts and community hospital endoscopists were observed. These findings strengthen the value of the Prague C&M classification to describe Barrett's segment length in clinical practice.

Inter-observer agreement (i.e. ICC) of C&M Barrett's length during real-time endoscopy was found to be almost perfect (ICC 0.91-0.92) and it was moderate for hiatal hernia length (ICC 0.59). These results are surprisingly similar with previous studies showing ICC 0.92-0.94 for C&M length and ICC 0.44 for hiatal hernia length. In these previous studies, however, video sequences
Part I: Imaging in Barrett’s esophagus

were scored either by endoscopists with a special interest in the Barrett’s esophagus or by Asian endoscopists.\textsuperscript{10, 15} Our study included Western endoscopists with and without special interest in Barrett’s esophagus and was performed during real-time consecutive endoscopies at different endoscopic sessions. The assessing endoscopists were also the ones who actually performed the endoscopy, controlling the amount of air insufflation during endoscopy - which is especially important when defining the position of the upper end of the gastric folds - and deciding on the moment to read off the location of a landmark, while advancing or retrieving the endoscope. Furthermore, the assessment of the location of the landmarks was based on the 5 cm interval markings on the endoscope, instead of the 1 cm intervals used in video sequences.

Given the above mentioned factors, we had expected to find a lower inter-observer agreement for the C&M classification compared to the standardized evaluation of video sequences. Although these results seem to support the validity of the C&M criteria, one can argue that the ICC may not be a correct method to assess inter-observer agreement of the C&M classification as mentioned in the methods section. The ICC value in our study may be artificially high due to the wide range of Barrett’s length in our patients (range C length 0-18 cm, range M length 1-18 cm).

The hypothesis that the ICC in our study may overestimate the agreement of the C&M length is supported by the absolute agreement, which was low for C&M length when 0 cm discrepancy was allowed. In only 31% (for C length) and 39% (for M length) of the patients the two endoscopists completely agreed on C&M length. The absolute agreement increased to 68% and 74% when a discrepancy up to 1 cm was allowed and to 89% and 91% for a discrepancy ≤2 cm. These percentages are only slightly lower than those found in the original validation study using video sequence scoring: 38 and 53% for 0 cm discrepancy, 82 and 88% for ≤1 cm discrepancy, and 95 and 97% for ≤2 cm discrepancy.\textsuperscript{10} Absolute agreement during real-time endoscopy is thus high but we should anticipate that C&M values length may vary by 1-2 cm between two endoscopies. From a clinical perspective, a difference of 1-2 cm between two endoscopists may be less relevant when evaluating long Barrett’s segments than shorter segments. The limits of agreement of the Bland-Altman plots do not increase with increasing length of the variable (C, M, or hiatal length). In addition, absolute agreement is not related to the length of the Barrett’s segment. We therefore also created the relative agreement by dividing the concordance between the two consecutive endoscopies by the mean length of the variable of interest (C length, M length and hiatal hernia length) for each patient. In accordance with the inter-observer agreement and absolute agreement, the relative agreement was also high for C&M length (91%) and slightly lower for hiatal hernia length (80%). Previous studies have shown that hiatal hernia length varies considerably over time.\textsuperscript{16} This phenomenon might account for the relatively lower agreement in hiatal hernia length found in our study. Additionally, respiratory movements and the amount of insufflation have also an impact on the length of the measured hiatal hernia length.
When comparing Barrett’s segments $M \leq 5$ cm with longer Barrett’s segments ($M > 5$ cm) we found ambiguous results: Barrett’s $M > 5$ cm had a significantly higher ICC but a lower absolute agreement (Table 4 and 5). The relatively higher ICC value probably reflects the difference in variance of Barrett’s length in both groups as explained above: a higher ICC was found for Barrett’s $M > 5$ cm which had a wider variance (M6-18 cm), while a lower ICC was found for Barrett’s $M \leq 5$ cm which had a smaller variance (M1-5 cm). The relative agreement, however, supports the ICC value as it shows also a significantly higher relative agreement in M length for Barrett’s $M > 5$ cm. Therefore it seems that endoscopists agree less on C&M length in shorter Barrett’s segments.

When comparing Barrett’s experts and community hospital endoscopists, no differences were found in inter-observer agreement, absolute agreement or relative agreement despite the differences in patient population between both groups. In group I patients were included with Barrett’s C or M >2 cm, while group II consisted of patients with any length of Barrett’s. This may have influenced the ICC of the community hospital endoscopists in two ways: ICC of community hospital may have been overestimated, as group II had a wider range in Barrett’s length; or ICC may have been underestimated, as small Barrett’s segments have shown to have a lower inter-observer agreement.10, 15 Nevertheless, our results suggest that endoscopists without specific Barrett’s expertise are also able to use the C&M classification adequately.

This study is not without limitations. First, as mentioned above, the variance of Barrett’s length in the study population impacts the calculated values of the ICC. Second, the patient populations evaluated by Barrett’s experts and community hospital endoscopists were not completely comparable. Third, the low number of patients with short Barrett’s segments: this study only included 3 patients with a Barrett’s segment <1 cm. This was partly caused by the inclusion criterion of Barrett’s of C or M >2 cm in group I. Another possible contributing factor might be that including neoplastic Barrett’s segments (LGIN or HGIN and/or cancer) may have resulted in longer Barrett’s segments as the length is known to correlate with risk of neoplasia. Another potential limitation are possible systemic differences. Although no systemic difference were found this might be caused by a relatively low number of patients. Finally, we can not exclude that especially community hospital endoscopists have been more careful with the measurements than normally in clinical practice as it concerned clinical trial. Nevertheless, endoscopists were not aware during the procedures that the C&M landmarks would be used for inter-observer evaluation as this was a post hoc analysis.

This study had also several strengths. First, endoscopies were performed consecutively at separate endoscopic sessions and were assessed independently by endoscopists blinded to the previous C&M values. Second, endoscopies were performed under standardized conditions with same case
record forms, sedation and endoscopic circumstances. Third, this study included homogenous
cohorts of endoscopists utilizing similar endoscopic expertise at the two corresponding
endoscopies. Fourth, Barrett's expert endoscopists were compared to community hospital
endoscopists, although no direct comparison was possible as Barrett's experts and community
hospital endoscopists assessed a different group of patients. Finally, endoscopic assessment of the
C&M landmarks was performed in real-time instead on video sequences.

Our results shows that the Prague C&M classification has a good agreement during real-time
endoscopy when applied by Barrett's expert and community hospital endoscopists, indicating
its usefulness in clinical practice, not only in tertiary referral centers with extensive expertise
in Barrett's imaging, but also among community endoscopists. In our opinion, assessment and
registration of the Barrett's length incite the endoscopist to inspect and observe more carefully
the Barrett's segment. In addition, knowing the Barrett's length is important when performing
endoscopic therapy as it is necessary to plan and evaluate the therapy as well. In the future,
assessment and registration of Barrett's length might be helpful for cancer risk stratification,
e.g. by lengthening or shortening surveillance intervals according to the length of the Barrett's
segments. In this respect it is important to stress our finding that absolute agreement in BE length
was only high when we accepted a difference of ≤2 cm.

In conclusion, application of the Prague C&M classification during real-time endoscopy
showed good agreement for assessing circumferential (C) and maximum (M) Barrett's length
and reasonable agreement for assessing hiatal hernia length. Agreement for shorter Barrett's
segments (M≤5 cm) seems to be lower. No differences in agreement were observed between
Barrett's experts and community hospital endoscopists. These findings strengthen the value of
the Prague C&M classification in clinical practice to describe the length of the Barrett's segment.

Acknowledgements

The authors thank LC Baak, C Böhmer, AH Naber, CY Ponsioen, MB Wallace, KK Wang, and V
Subramanian for their participation as endoscopists.
References


