The Development and Validation of an Endoscopic Grading System for Barrett's Esophagus: The Prague C & M Criteria

PRATEEK SHARMA,* JOHN DENT,‡ DAVID ARMSTRONG,§ JACQUES J. G. H. M. BERGMAN,¶ LIEBWIN GOSSNER,‖ YOSHIO HOSHIHARA,# JANUSZ A. JANKOWSKI,** OLA JUNGHARD,‡‡ LARS LUNDELL,§§ GUIDO N. J. TYTGAT,§ and MICHAEL VIETH¶¶

**Department of Veterans Affairs Medical Center, University of Kansas School of Medicine, Kansas City, Missouri; ‡Department of Gastroenterology, Hepatology, and General Medicine, Royal Adelaide Hospital, Adelaide, Australia; §Division of Gastroenterology, McMaster University, Hamilton, Ontario, Canada; ¶Department of Gastroenterology and Hepatology, Academic Medical Centre, University of Amsterdam, The Netherlands; ‖Second Department of Internal Medicine, HSK-Kliniken, Wiesbaden, Germany; #Department of Gastroenterology, Toranomon Hospital, Tokyo, Japan; **Department of Clinical Pharmacology, University of Oxford, Oxford, United Kingdom; ‡‡AstraZeneca R&D, Mölndal, Sweden; §§Division of Surgery, Karolinska University Hospital, Stockholm, Sweden; and ¶¶Institute of Pathology, Otto-von-Guericke University, Magdeburg, Germany*

See CME Quiz on page 1626.

Background & Aims: Barrett's esophagus (BE) is a premalignant condition for esophageal adenocarcinoma, its diagnosis relying initially on recognition of a columnar-lined distal esophagus. We aimed to develop and validate explicit, consensus-driven criteria for the endoscopic diagnosis and grading of BE.

Methods: An international working group agreed on criteria and developed materials for their formal evaluation using video-endoscopic recordings gathered in a standardized manner in 29 patients. The criteria included assessment of the circumferential (C) and maximum (M) extent of the endoscopically visualized BE segment as well as endoscopic landmarks. The recordings were scored according to these criteria by a separate international panel of 29 endoscopists. **Results:** The Prague C & M Criteria give explicit guidance on the endoscopic recognition of BE and grading of its extent. The overall reliability coefficients (RC) for the assessment of the C & M extent of the endoscopic BE segment above the gastroesophageal junction were 0.95 and 0.94, respectively. The rates of exact agreement (for C & M values) for pairwise comparisons of individual patient values were 53% and 38%, respectively, whereas the values for agreement within a 2-cm interval were 97% and 95%, respectively. The overall RC for endoscopic recognition of BE $\geq$1 cm was 0.72, whereas for BE $<$1 cm, it was 0.22. The RCs for recognizing the location of the gastroesophageal junction and the diaphragmatic hiatus were 0.88 and 0.85, respectively. **Conclusions:** The Prague C & M Criteria have high overall validity for the endoscopic assessment of visualized BE lengths.

Barrett's esophagus (BE) is the premalignant condition for adenocarcinoma of the esophagus and gastroesophageal junction (GEJ) and is associated with chronic gastroesophageal reflux disease.[1,2] In BE patients, the proximal level of the squamocolumnar junction (SCJ or Z-line) is such that it no longer coincides with the GEJ. The resulting columnar-lined mucosa of the distal esophagus appears salmon-pink in color and is readily visible by endoscopic examination. Biopsy specimens can then be obtained from this area of suspected BE (endoscopic BE) to characterize further the tissue and to document specifically the intestinal metaplasia.[3,4] Thus, reliable diagnosis of BE, with its associated risk for esophageal adenocarcinoma, depends first on the effective recognition of endoscopic features of the suspected BE segment, followed by technically adequate histologic sampling of the metaplastic esophageal mucosa to screen for intestinal-type metaplasia. Apart from biopsy sampling, measurement of the extent of BE is also clinically relevant because this influences the risk of developing adenocarcinoma.[5,6] For instance, the Rotterdam Esophageal Tumour Study Group found that a doubling of the length of BE increased the risk of adenocarcinoma by 1.7 times.[6]

Despite the importance of accurate endoscopic recognition and grading of BE, there is no consensus-based, authoritative guidance on how this should be done. Accordingly, a variety of ad hoc and frequently inadequately specified and validated approaches have been used. For instance, grading of patients into those with variably defined "short" and "long" segments of BE is an unsatisfactorily crude approach.[7] Previous studies have found considerable inter- and intraobserver variation, even when an endoscopic length of BE of more than 3 cm is estimated.[8,9] Clinical management and research of BE patients would benefit greatly from the widespread use of standardized, simple, and practical criteria for the endoscopic grading of BE with standard endoscopic equipment. This report describes the consensus-based development and subsequent validation of criteria designed to achieve this aim. The major specifications for these criteria were that they should give the best possible guidance for accurate endoscopic recognition of the anatomical GEJ (the dividing line between the esophagus and stomach) as the first and crucial step in the recognition and then grading of BE. Furthermore, the approach to definition of extent has to be able to capture the great variability of orad circumferential and tongue-like extents of BE and to express them in a way that is easily understood. The criteria that have been developed have been named the Prague C & M Criteria because they were first presented at an open meeting at the Prague September 2004 United European Gastroenterology Week.

Abbreviations used in this paper: BE, Barrett's esophagus; C, circumferential; GEJ, gastroesophageal junction; M, maximum; SCJ, squamocolumnar junction.

0016-5085/06/$32.00
doi:10.1053/j.gastro.2006.08.032

Throughout this report, the term "Barrett's esophagus" is used to indicate the macroscopic identification, by standard endoscopy, of abnormal columnar esophageal epithelium suggestive of columnar-lined distal esophagus. This is the first step in the recognition of patients with this lesion, before biopsy specimens are obtained.

Materials and Methods

Working Group Structure and Meetings

A subgroup of the International Working Group for the Classification of Oesophagitis (IWGCO), the authors of this report, was assembled to develop criteria for endoscopic detection and grading of BE that would be useful in both clinical practice and research trials. All the members of the subgroup have a research and clinical interest in BE, and one is a statistician. The first meeting was convened in September 2002, and, here, strategies for development of criteria for grading of BE were formally considered. Statements about the recognition of key endoscopic landmarks pertinent to the endoscopic recognition and classification of BE were developed and rated by the group. The level of evidence in support of each statement was provided, and the members voted anonymously regarding their acceptance or refusal of each statement, according to the voting structure.[10] Voting and discussion on the statements helped identify issues that needed further consideration. Subsequent meetings and work by individual group members in preparation for these meetings addressed these outstanding issues.

Throughout the development of the endoscopic criteria, video recordings of patients with and without endoscopic evidence of BE were used at the group's meetings to test and refine approaches toward the recognition of endoscopic landmarks and for implementation of criteria. These recordings of endoscopic examinations were made by some members of the group and/or their colleagues (see Acknowledgements section). Endoscopes with the best possible image resolution were used, and digital video recorders captured the images. The digital format allowed subsequent copying of recordings without loss of image quality.

Preparation of Videos of Endoscopy Procedures

The group had agreed at the outset of its work that the final criteria should be validated by a study of interobserver agreement. The positions of landmarks used in the criteria were to be scored in video recordings of endoscopies in patients with varying lengths of BE and in those without any endoscopic evidence of BE. To achieve this goal, the group developed a standard approach for making each video recording that was judged to best enable recognition of the positions of landmarks, and a library of 50 such video recordings was collected. The method of endoscopic recordings avoided giving any cue to the observer about the display of landmarks relevant to judgments on the presence and extent of BE.

The protocol used to make the video recordings was defined in detail, in writing, and had the following major features. The depth of endoscope insertion was monitored and recorded in a standardized manner by an assistant who spoke the depth of endoscope insertion (as judged by the centimeter markings at the bite block) every few seconds. The assistant's documentation of endoscope insertion depth was audio recorded directly onto the digital videotape during image acquisition. Video endoscopic images were gathered at each centimeter of depth of endoscope insertion with the endoscope being maintained at each level long enough to display all findings. The endoscopist then moved the endoscope gradually, with the assistant saying "moving" as this was done and also providing the depth of insertion to which the endoscope had been moved. Two sequences of such 1-cm stepwise insertion and withdrawal were recorded over the full distance required to demonstrate clearly the findings between the uppermost margin of the squamocolumnar junction (SCJ) and the cavity of the upper stomach, clearly below the diaphragmatic hiatal impression if a hiatus hernia was present. During video recordings, endoscopists were instructed to maintain air insufflations sufficient to provide good visualization of the landmarks and mucosa but insufficient to efface the upper ends of the gastric mucosal folds.

The video/audio recordings were sent to a central editing and storage facility (QPC, Gothenburg, Sweden) at which they were coded and then anonymized by deletion of all printed information recorded with the endoscopic images. The editing process included transfer of the audio documentation of depth of endoscope insertion into a numeric display of this value onto the video recording above the top left-hand corner of the video image.

Internal and External Validation Study

Two validation studies were carried out: the initial study was conducted with members of the IWGCO working group as assessors (internal study), and the second study was conducted with external assessors who had not participated in the recording and selection of the video sequences (external study). An initial video sequence-based validation study was done on 50 video clips made during both endoscope insertion and withdrawal. Videos were scored according to a standard form developed by the group. The data and experience from this preliminary internal study led to the final choice of 29 video sequences, gathered only during endoscope withdrawal. Experience with the video clip scoring form led to its significant revision so that its clarity was substantially improved for the definitive study.

For the study with external assessors, members of the working group recruited 29 expert endoscopists with a special interest in BE (see Acknowledgements section) from 14 countries to evaluate each of the 29 video sequences for the standardized criteria presented in results, using a standard scoring form (Figure 1). The video sequences were arranged in random order, assigned code numbers, and copied onto a digital video disc (DVD). The DVD also contained an instructional video on how to score the developed criteria, which had to be watched before any of the test videos could be evaluated. This instructional video used a high-quality endoscopic withdrawal sequence from a patient with a hiatus hernia and a BE length >3 cm. Instructions were provided explicitly regarding the measurement of contiguous segments of BE only, ie, islands of squamous and columnar mucosa were not to be included in the assessment. Landmarks were highlighted by the use of freeze-framing and superimposition of lines and arrows edited onto the image, with accompanying explanatory text beside the video image. This instructional video also explained to the assessors how they could control the playing of video sequences, including freeze-framing and slow motion scrolling through the se-

1. Could you identify the diaphragmatic hiatus? ○ yes ○ no
   If yes, what was the endoscope insertion distance at the diaphragmatic hiatus? ☐ cm

2. Could you identify the proximal margin of the gastric fold? ○ yes ○ no
   If yes, what was the endoscope insertion distance at the proximal ends of the gastric folds? ☐ cm

3. Could you identify a circumferential 'ring' or 'pinch' near the proximal end of the gastric folds? ○ yes ○ no
   If yes, what was the endoscope insertion distance at this point? ☐ cm

4. Was there a circumferential segment of Barrett's esophagus? ○ yes ○ no
   If yes, what was the endoscope insertion distance at the proximal margin of the circumferential Barrett's esophagus? ☐ cm

5. Was there at least one tongue-like segment of Barrett's esophagus? ○ yes ○ no
   If yes, what was the endoscope insertion distance at the proximal margin of the longest Barrett's tongue a this point? ☐ cm
   What portion of the circumference of the esophagus did the tongue occupy? ○ <50% ○ ≥50%

6. With respect to the extent of the Barrett's esophagus, how would you classify this according to the length from the gastroesophageal junction?
   Circumferential (C) segment? C ☐ cm
   Maximum extent (M), including tongues? M ☐ cm

7. Were there any mucosal breaks identified in the squamous mucosa? ○ yes ○ no
   If yes, what was the severity of the esophagitis? **Please check (✓) one grade.**
   Los Angeles (LA) ○ Grade A ○ Grade B ○ Grade C ○ Grade D

8. Technical adequacy of images. **Please check (✓) one number per clip.**
   ○ Not scorable ○ Poor ○ Fair ○ Good ○ Excellent

Figure 1. Assessment score sheet for analysis of endoscopy videos.

quences. The DVD also contained detailed text-based instructions (Figure 2) and the score form. The form was completed electronically and submitted to a Web site set up for the validation study. Assessors had approximately 4 weeks to complete their evaluation, at which time the database was closed for analysis.

Video Clip Assessment and Data Capture

The recordings were assessed using a hybrid DVD/Web solution via a dedicated Web site on a structured query language (SQL) database in conjunction with a specially prepared DVD containing the instructional video and the clips to be assessed. Assessors were given a set of instructions and were asked to identify key landmarks from each clip. Their observations were marked on the on-line score sheet, which was entered directly into the remote database.

Statistical Analysis

The level of evidence in support of each statement was provided, and the members voted anonymously regarding their acceptance or refusal of each statement, according to the voting structure. Consensus was defined as greater than 70% of votes cast in agreement or disagreement with each proposition.

Interrater agreement regarding the length of C & M was evaluated by the intraclass correlation coefficient (or interrater reliability coefficient). With 2 raters, the intraclass correlation coefficient is the same as the weighted κ with quadratic weights, and, with more than 2 raters, it approximates the mean of all pairwise weighted κ values. As another measure of interrater agreement regarding the length of C & M, the proportion of pairwise comparisons in which the raters' assessments agreed exactly or differed by at most 1 cm and at most 2 cm, respectively, was calculated. Interrater agreement regarding binary variables (eg, presence/absence of BE) was also evaluated by the intraclass correlation coefficient, in this case approximating the average of all pairwise ordinary κ values.

The strength of rater agreement was categorized according to definitions proposed by Landis and Koch for κ values.[11] These were as follows: 0.00–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–1.00, almost perfect. The sample size (comprising the number of raters and video clips) for the external validation was chosen to yield a difference of approximately 0.05 (at most 0.1) between the estimate and the

General comments

For all landmarks, record the depth of endoscope insertion in centimeters *(as shown on screen)*, at the point where the feature is observed, on the data form as follows

- The depth of insertion of the endoscope for a feature is determined as the point just before the feature comes into full view on withdrawing the endoscope
- For any length of endoscopic Barrett's epithelium <1 cm above the GEJ, please report the measurement as "<1 cm"
- For all other lengths of endoscopic Barrett's epithelium, and for other landmarks, please report the value rounded to the nearest centimeter *(i.e. rounded down if <5 mm and rounded up if ≥5 mm)*

Criteria for GEJ *(features to be sought on video)*

- Proximal extent/tips of gastric folds, or pinch ring
- Exact measurements of the depth of endoscope insertion should be made and reported on the data sheet
- The GEJ location should be determined by either:
  - the most proximal extent of the gastric folds, *or*
  - the circumferential 'ring' or 'pinch' just above to the proximal ends of the gastric folds (indicating the distal end of the lower esophageal sphincter)

Criteria for limits of Barrett's epithelium *(features to be sought on video)*

- If there is any endoscopic suspicion of Barrett's epithelium, the extent of the 'circumferential' and any proximal 'tongue-like' changes should be determined
- Exact measurements of endoscope insertion depth should be made and reported on the data sheet for:
  - the proximal margin of the circumferential segment of Barrett's epithelium
  - the proximal margin of the longest 'tongue-like' segment of Barrett's epithelium

Criteria for reporting the extent of endoscopic Barrett's esophagus

The 'C' (circumferential) and 'M' (maximal) measurements are calculated, respectively, as:

C = The difference in endoscope insertion distances between the positions recorded for the "GEJ" and the "proximal margin of the circumferential Barrett's epithelium"

M = The difference in endoscope insertion distances between the positions recorded for the "GEJ" and the "proximal margin of the longest 'tongue-like' segment of Barrett's epithelium" *(do not include Barrett's islands in this assessment)*

For example:

C3M5: Circumferential Barrett's at 3 cm above GEJ and a maximum extent of 5 cm *(tongue extending 5 cm above GEJ)*

C0M3: No circumferential extent of Barrett's and a maximum extent of 3 cm *(tongue extending 3 cm above GEJ)*

C2M2: Circumferential Barrett's at 2 cm above GEJ and a maximum extent of 2 cm *(no tongues extending beyond circumferential area of Barrett's esophagus)*

C<1M<1: Circumferential Barrett's at <1 cm above GEJ and a maximum extent <1 cm above the GEJ *(tongue extending <1 cm above GEJ. This is equivalent to an irregular Z-line)*

Figure 2. Endoscopic grading system for Barrett's esophagus. A copy of these instructions was given to assessors participating in the validation of the grading system. *GEJ*: gastroesophageal junction.

Table 1. Key Propositions and Acceptance Levels Among the Barrett's Esophagus Working Group Participants

Proposition and level of acceptance	Acceptance level (%)[a]
"Barrett's esophagus," qualified by a descriptor of its extent, is the most pragmatic term to describe the endoscopic appearances suggestive of columnar-lined esophagus.	1 (100)[b]
Measurement of the nearest and farthest proximal limits of Barrett's esophagus, relative to the distal limit (the GEJ), is a reasonable assessment of its extent.	1 (100)[b]
The proximal limit of linear gastric mucosal folds is the most practicable indicator of the gastroesophageal junction (GEJ) in the presence of suspected Barrett's esophagus in routine diagnostic endoscopic practice.	1 (83)[b] 2 (17)[b]
The proximal limit of gastric mucosal folds is defined best as the most proximal point at which there is any evidence of a linear fold of gastric mucosa. This is best visualized when the esophagus is distended minimally to the point that the proximal ends of the gastric folds appear.	1 (100)[b]
Endoscopic changes suggestive of columnar-lined esophagus, extending less than 1 cm above the GEJ, are of uncertain value for the diagnosis of Barrett's esophagus.	1 (100)[b]
Endoscopic changes suggestive of columnar-lined esophagus, extending at least 1 cm above the GEJ, are likely to predict the presence of Barrett's esophagus.	1 (83)[b] 2 (17)[b]
Assessment of the extent of Barrett's esophagus should be conducted by measuring the depth of insertion of the endoscope relative to the bite block.	1 (100)[b]

[a]Workshop participants voted on whether they accepted the proposition completely (1), with some reservations (2), or with major reservations (3), or rejected the propositions with reservations (4), or completely (5).
[b]Greater than 70% of votes cast in agreement or disagreement of each proposition was established as consensus.

lower limit for a 1-sided 95% confidence interval for intraclass correlation coefficients.

Results

Endoscopic Landmarks and Development of a Classification System

A series of statements about the recognition of key landmarks pertinent to the endoscopic recognition and classification of BE were developed and rated by the group (Table 1). This process highlighted the pivotal importance of accurate localization of the GEJ for the diagnosis of BE and also for the measurement of its extent. Review of the literature relevant to the endoscopic location of the GEJ showed a lack of validation of any of the criteria proposed. None of these criteria were originally proposed by the working group. Because of the lack of authoritative guidance in the literature on how to locate the GEJ, the working group judged it necessary to evaluate and describe how best to assess the position of the GEJ endoscopically. Of the 3 options considered at the initial meeting of the working group, the only one supported by all members was the GEJ as being "at the proximal margin of the gastric mucosal folds." This support was given despite acceptance by the group that the reliability of determining this landmark was dependent on minimization of the distending volume of air used during endoscopy (Table 1). The second option for location of the GEJ was "a pinching of the lumen by the narrow band of the specialized muscle of the lower esophageal sphincter," ie, the "sphincteric pinch." This was considered worthy of evaluation but was less strongly supported than the proximal margin of the gastric mucosal folds. Use of the lowest extent of the palisade blood vessels of the esophagus as a marker of the GEJ was also discussed at length. The group's overall opinion of the utility of this criterion was, however, lukewarm. This was on the basis that discrimination between the true palisade vessels and other vascular patterns present in mucosa below the GEJ was sometimes difficult and that the palisade vessels could not be seen adequately in all patients when standard endoscopic imaging methods were used.

The final criteria agreed on by the working group are given in Figures 3 and 4. They were refined at multiple meetings of the group, and their last evolution occurred just before the launching of the definitive validation study by external assessors.

Development of the C and M Criteria

The relatively simple approach of recording both the circumferential extent (the C value) and the maximum extent (the M value) above the GEJ in centimeters (Figure 3) proved to be easy to use when tested by the group and was judged to give the best balance between simplicity and capture of potentially useful detailed information on extent. It was agreed that true islands of squamous and columnar mucosa should not influence the measurement of extent and that this should be stated in the subtext for the criteria. The choice of "M" for "Maximum extent" was a much debated late evolution of the criteria from the previous use of "T" for "Total extent." This change was made because the internal preliminary validation study indicated the potential for "T" to be mistakenly interpreted as the

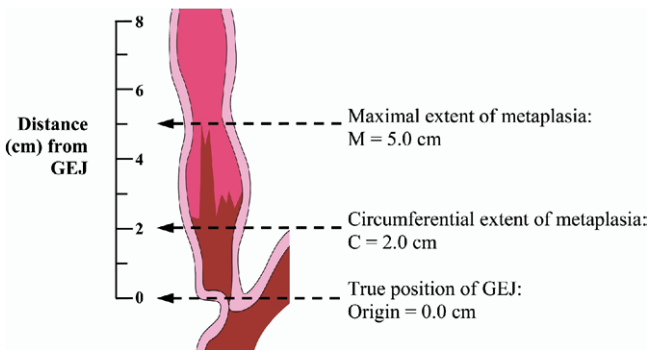

Figure 3. Diagrammatic representation of endoscopic Barrett's esophagus showing an area classified as *C2M5*. *C*: extent of circumferential metaplasia; *M*: maximal extent of the metaplasia (C plus a distal "tongue" of 3 cm); *GEJ*: gastroesophageal junction.

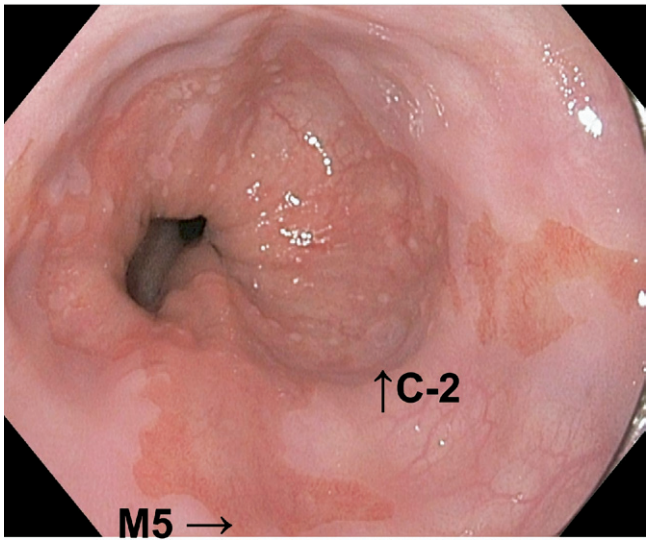

Figure 4. Video still of endoscopic Barrett's esophagus showing an area classified as *C2M5*. *C*: extent of circumferential metaplasia; *M*: maximal extent of the metaplasia (C plus a distal "tongue" of 3 cm).

length of "tongues" of BE, rather than the total length above the GEJ.

Thus, the grading system defined by the working group to improve the recognition of and reporting of gastroesophageal landmarks and endoscopically recognized BE included the C & M extent of endoscopically recognized BE, GEJ, SCJ, and diaphragmatic hiatus (Figure 2). Figures 3 and 4 show the C & M extents of endoscopically recognized BE, with C = 2 cm and M = 5 cm, giving a classification of C2M5.

Initial Validation of the Classification System: Internal Study

The grading system was validated initially by a panel of 5 members of the working group, who assessed a selection of 50 video clips. The video clips were viewed in random order. The internal assessment produced reliability coefficients of 0.91 for C and 0.66 for M. This correlates to an "almost perfect" level of reliability for C and "substantial" reliability for M (Table 2). One assessor misinterpreted M as being the "tongue" length, and, if the results from this assessor were excluded, the reliability coefficients were 0.94 for C and 0.88 for M. There were only minimal differences between the reliability coefficients for push-only and pull-only endoscopic procedures (Table 2), indicating that these criteria could be used either during endoscope insertion or toward the completion of endoscopic procedure, ie, withdrawal.

Table 2. Reliability Coefficients for the Initial Validation of the Classification System: Internal Study

	All endoscopies (push or pull)	Push-only endoscopy	Pull-only endoscopy
Circumferential extent (C)	0.91 (0.94)[a]	0.93 (0.94)[a]	0.91 (0.94)[a]
Maximal extent (M)	0.66 (0.88)[a]	0.65 (0.96)[a]	0.67 (0.81)[a]

[a]Reliability coefficient if the results from 1 of the 5 internal assessors, who did not understand the "M" classification, are not included in the analysis.

Table 3. Number of Video Clips With C & M Assessments in Relationship to the Length of the BE Segment

Estimated BE length	Number of video clips (C value)	Number of video clips (M value)
0.0 to <0.5 cm	14	5
0.5 to <1.0 cm	4	2
1.0 to <3.0 cm	4	11
3.0 to <5.0 cm	2	4
≥5.0 cm	5	7

Validation of the Classification System: External Study

Of the 29 external assessors invited to participate in the analysis, 22 submitted complete data for C & M values for the selection of the 29 video clips selected for this study. One observer assessed only 1 video clip, and these data were excluded from analysis. Moreover, 9 observers had at least once recorded an M value that was numerically smaller than the C value on the same clip (the M value should always be ≥C value). In these situations, the M value was replaced with the C value. The distribution of mean C & M assessments of the 29 video clips is presented in Table 3. Almost half of the C assessments but only 5 of the M assessments were less than 0.5 cm.

The overall reliability coefficients from the external assessment were 0.94 for C and 0.93 for M, representing an "almost perfect" level of reliability for both. Using the C & M criteria, assessors were able to agree on the presence of endoscopic BE greater than 1 cm in length with substantial reliability (RC = 0.72). The recognition of endoscopic BE <1 cm in length was only slightly reliable (RC = 0.21), making the recognition of endoscopic BE of any length moderately reliable (RC = 0.49). The assessors were able to recognize the proximal margin of the gastric folds and the diaphragmatic hiatus with almost perfect reliability (RC = 0.88 and 0.85, respectively). When calculating percentage agreement, each observer was compared with every other observer. For such pairwise assessment, there were a total of 6699 comparisons from the 29 video clips. Of these comparisons for C & M values, the exact rates of agreement were 53% and 38%, respectively. The comparisons differed at most by 1 cm in 88% and 82% and differed at most by 2 cm in 97% and 95% of the C & M values, respectively. The detailed breakdown of results from the external assessment by length of BE and reliability coefficients for recognizing the position of gastroesophageal landmarks are presented in Tables 4–6.

There were no observers that recorded extreme values, ie, consistently the highest or lowest recordings. The observer with the highest number of extreme recordings had, out of the 29 clips, 3 highest recordings on C and 4 highest recordings on M. The results did not change when this observer was excluded from the analysis.

Discussion

At present, standardized, validated criteria for the endoscopic description of BE are not routinely used. Endoscopists currently adopt a loose classification system, defining endoscopic segments of BE as "long," "short," or "ultra-short," with-

Table 4. Reliability Coefficients for Recognizing Different Lengths of Barrett's Esophagus and the Position of Esophageal Landmarks

Recognition of BE or position of gastroesophageal landmark	Reliability coefficient	Reliability (Landis and Koch, 1977)[11]	Lower 2-sided 95% CI	Upper 2-sided 95% CI
Diaphragmatic hiatus	0.85	Almost perfect reliability	0.78	0.91
Proximal margin of gastric fold	0.88	Almost perfect reliability	0.82	0.93
Pinch at distal esophagus	0.78	Substantial reliability	0.68	0.87
C value	0.94	Almost perfect reliability	0.91	0.97
M value	0.93	Almost perfect reliability	0.89	0.96

CI, confidence interval.

out there being an established cutoff or clinical significance for any of these categories.[7] Moreover, considerable variation in the ability to detect, classify, and measure the endoscopic extent of BE has been reported.[8,9]

To deal with these issues, we developed an endoscopic classification system for BE that would be useful for clinical practice and in clinical trials and that would act as a foundation for subsequent histologic analysis. The endoscopic recognition of BE requires identification of endoscopic landmarks, namely, the GEJ, squamo-columnar junction, and diaphragmatic hiatus. These landmarks document the presence of a columnar-lined esophagus (ie, esophageal columnar metaplasia). Further characterization of BE requires multiple biopsy specimens from the columnar-lined segment to be obtained to detect metaplastic and dysplastic epithelium. The Prague C & M Criteria alert the endoscopists to the presence of a columnar-lined esophagus and thereafter allow the grading of the C & M extent of the columnar-lined esophagus (BE segment). It was not the intent of the group to exclude histology as a vital tool for assessment of patients with BE. Rather, we sought to focus on the endoscopic recognition and grading of endoscopic criteria, which is the initial crucial step prior to obtaining correctly located biopsy specimens. Also, it was agreed that true islands of squamous and columnar mucosa should not influence the measurement of extent of BE and that only segments of contiguous BE be measured. This study is the first to grade the endoscopic extent of BE in a standardized manner and using consensus-based and validated criteria. It is also the first to validate formally the identification and measurement of endoscopic landmarks, although endoscopists have been using these features for some time.

Endoscopists participating in the external validation reliably recognized the position of esophageal landmarks such as the proximal margin of the gastric folds, pinch at the distal esophagus, and diaphragmatic hiatus (giving reliability coefficients of 0.88, 0.78, and 0.85, respectively). This demonstrates a substantial-to-near-perfect level of reliability and suggests that there is little difference between expert endoscopists in locating these key landmarks. Recognition of the GEJ is of crucial importance for the reliable and consistent detection of BE at endoscopy. We evaluated the proximal margin of the gastric folds as a marker of the GEJ.[12] It was acknowledged by the group that a definition of the most distal ends of the palisade-shaped longitudinal vessels may be of value, based on the methods of identifying the GEJ in Japan.[13] However, the palisade vessels may not be uniformly visible using standard endoscopy in reflux patients and may be obscured by the presence of a double muscularis mucosa in patients with BE.[13,14] Some centers use differences between mucosal surface patterns to identify the GEJ, but these generally require chromoendoscopy or high-resolution endoscopy, which still are not currently practicable for routine use. Gastric folds have the advantage of being independent of the BE length, hiatus hernia, or changes in the gastric mucosa and were therefore judged to be the most practical and most commonly used landmark of the GEJ. Consensus was also obtained for a very similar proposition at the AGA Chicago workshop.[10] Furthermore, all members agreed that the proximal limit of gastric mucosal folds are best visualized when the esophagus is distended minimally, although further work may be required to define "minimum insufflation."

The high reliability coefficient values for the C & M extent of endoscopically recognized BE (C & M values), particularly those obtained from the external study, suggest that endoscopists found the new terminology easy to understand and were able to measure C & M from video clips of BE with a high degree of reliability. Endoscopists were not able to reliably measure shorter lengths of BE (<1 cm). This is not surprising because there is a lower margin for error when identifying the GEJ and extent of endoscopically recognized BE when the segment is very short.

Despite the impressive reliability of the Prague C & M classification system, this study has a number of potential limitations. All assessments were performed using video clips rather than actual patients. However, it would be logistically impossible to base such a study on live endoscopy. The use of video clips allowed us to gather data from a large number of endoscopists. The clips were screened for how adequately

Table 5. Percentage Agreement for C & M Values

	Percentage agreement
Exact agreement: C	53
1-cm difference: C	88
2-cm difference: C	97
Exact agreement: M	38
1-cm difference: M	82
2-cm difference: M	95

Table 6. Overall κ Values, Based on BE Length

	κ Values
BE, any length	0.49
BE, length ≥1 cm	0.72
BE, length <1 cm	0.21

they displayed endoscopic landmarks. Use of digital recording equipment ensured that the quality of stored images was equivalent to live endoscopy and the use of freeze-frame gave assessors the opportunity to analyze each clip in detail. The classification system described here does not measure the surface area of the columnar-lined esophagus, which may be more important than just the endoscopic extent.[6,15–17] One technique to accurately and reproducibly measure the surface area of columnar-lined esophagus uses a computer program to create a 2-dimensional map of the esophagus.[18] Such an approach is not, however, suitable for everyday practice. The C & M measurements provide a useful approximation of the extent of endoscopic BE and can also be performed relatively quickly. The classification system was validated by a panel of endoscopists with an interest in BE. This may have given rise to an artificially high level of reliability in identification of landmarks and recognition of BE because of increased awareness of these key features in this expert group. The reliability of these criteria when used by endoscopists who are not especially interested in BE or who are less experienced remains to be tested. Finally, any new classification system for BE would be further validated if linked to outcomes. Such an initiative is already underway, but it will be several years before data will be available.

It is important to recognize that no prospective randomized study has established that endoscopic screening and surveillance programs for BE decrease the rates of death from cancer, and decisions regarding the optimal interval for surveillance are based on assumptions made in an area in which few data are available.[19–21] Thus, screening and surveillance are still controversial issues among gastroenterologists.[10,21,22] The effectiveness of medical or surgical intervention in treating BE is also debatable.[23–25] Nevertheless, it is vital to develop a standardized method of measuring BE to assess the efficacy of treatments in individual patients, for the classification of patients in clinical trials, and for the development of treatment algorithms in clinical practice.

Future developments and research to improve further the endoscopic detection and classification of BE should also include simple measures such as the development of endoscopes with markings at 1 cm (or smaller) intervals to enable greater accuracy. Ultimately, it would be desirable to promote the more widespread use of a computer program to flatten 3-dimensional esophageal images, permitting an accurate calculation of the area of endoscopically recognized BE. Similar methods may allow for a better assessment of esophageal adenocarcinoma risk.

In conclusion, a consensus-based endoscopic classification system has been proposed and undergone extensive internal and external validation by trained endoscopists. This system, which determines the C & M criteria to describe and classify endoscopic BE, is simple and can be measured reliably by different endoscopists. The location of gastroesophageal landmarks is central to this classification, and the validation study described here has shown that these can be reliably identified and located by different endoscopists. Adoption of this standardized classification system may greatly enhance the ability of physicians to gauge the efficacy of treatments for BE in individual patients and the classification of patients with BE in clinical trials. In addition, it may help to define better the natural history of BE.

References

1. Mann NS, Tsai MF, Nair PK. Barrett's esophagus in patients with symptomatic reflux esophagitis. Am J Gastroenterol 1989;84:1494–1496.
2. Winters C Jr, Spurling TJ, Chobanian SJ, Curtis DJ, Esposito RL, Hacker JF III, Johnson DA, Cruess DF, Cotelingam JD, Gurney MS, et al. Barrett's esophagus. A prevalent, occult complication of gastroesophageal reflux disease. Gastroenterology 1987;92:118–124.
3. Jones TF, Sharma P, Daaboul B, Cherian R, Mayo M, Topalovski M, Weston AP. Yield of intestinal metaplasia in patients with suspected short-segment Barrett's esophagus (SSBE) on repeat endoscopy. Dig Dis Sci 2002;47:2108–2111.
4. Sharma VK, Chockalingam H, Hornung CA, Vasudeva R, Howden CW. Changing trends in esophageal cancer: a 15-year experience in a single center. Am J Gastroenterol 1998;93:702–705.
5. Gopal DV, Lieberman DA, Magaret N, Fennerty MB, Sampliner RE, Garewal HS, Falk GW, Faigel DO. Risk factors for dysplasia in patients with Barrett's esophagus (BE): results from a multicenter consortium. Dig Dis Sci 2003;48:1537–1541.
6. Menke-Pluymers MB, Hop WC, Dees J, van Blankenstein M, Tilanus HW. Risk factors for the development of an adenocarcinoma in columnar-lined (Barrett) esophagus. The Rotterdam Esophageal Tumor Study Group. Cancer 1993;72:1155–1158.
7. Sharma P, Morales TG, Sampliner RE. Short segment Barrett's esophagus—the need for standardization of the definition and of endoscopic criteria. Am J Gastroenterol 1998;93:1033–1036.
8. Dekel R, Wakelin DE, Wendel C, Green C, Sampliner RE, Garewal HS, Martinez P, Fass R. Progression or regression of Barrett's esophagus—is it all in the eye of the beholder? Am J Gastroenterol 2003;98:2612–2615.
9. Kim SL, Waring JP, Spechler SJ, Sampliner RE, Doos WG, Krol WF, Williford WO. Diagnostic inconsistencies in Barrett's esophagus. Department of Veterans Affairs Gastroesophageal Reflux Study Group. Gastroenterology 1994;107:945–949.
10. Sharma P, McQuaid K, Dent J, Fennerty MB, Sampliner R, Spechler S, Cameron A, Corley D, Falk G, Goldblum J, et al. A critical review of the diagnosis and management of Barrett's esophagus: the AGA Chicago Workshop. Gastroenterology 2004;127:310–330.
11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.
12. McClave SA, Boyce HW Jr, Gottfried MR. Early diagnosis of columnar-lined esophagus: a new endoscopic criterion. Gastrointest Endosc 1987;33:413–416.
13. Takubo K, Honma N, Aryal G, Sawabe M, Arai T, Tanaka Y, Mafune K, Iwakiri K. Is there a set of histologic changes that are invariably reflux associated? Arch Pathol Lab Med 2005;129:159–163.
14. Armstrong D. Review article: towards consistency in the endoscopic diagnosis of Barrett's oesophagus and columnar metaplasia. Aliment Pharmacol Ther 2004;20(Suppl 5):40–47.
15. Hamilton SR, Smith RR. The relationship between columnar epithelial dysplasia and invasive adenocarcinoma arising in Barrett's esophagus. Am J Clin Pathol 1987;87:301–312.
16. Avidan B, Sonnenberg A, Schnell TG, Chejfec G, Metz A, Sontag SJ. Hiatal hernia size, Barrett's length, and severity of acid reflux are all risk factors for esophageal adenocarcinoma. Am J Gastroenterol 2002;97:1930–1936.
17. Iftikhar SY, James PD, Steele RJ, Hardcastle JD, Atkinson M. Length of Barrett's oesophagus: an important factor in the development of dysplasia and adenocarcinoma. Gut 1992;33:1155–1158.
18. Kim R, Baggott BB, Rose S, Shar AO, Mallory DL, Lasky SS, Kressloff M, Faccenda LY, Reynolds JC. Quantitative endoscopy: precise computerized measurement of metaplastic epithelial surface area in Barrett's esophagus. Gastroenterology 1995;108:360–366.

19. Spechler SJ. Clinical practice. Barrett's esophagus. N Engl J Med 2002;346:836–842.
20. Conio M, Blanchi S, Lapertosa G, Ferraris R, Sablich R, Marchi S, D'Onofrio V, Lacchin T, Iaquinto G, Missale G, et al. Long-term endoscopic surveillance of patients with Barrett's esophagus. Incidence of dysplasia and adenocarcinoma: a prospective study. Am J Gastroenterol 2003;98:1931–1939.
21. Shaheen NJ, Crosby MA, Bozymski EM, Sandler RS. Is there publication bias in the reporting of cancer risk in Barrett's esophagus? Gastroenterology 2000;119:333–338.
22. Inadomi JM, Sampliner R, Lagergren J, Lieberman D, Fendrick AM, Vakil N. Screening and surveillance for Barrett's esophagus in high-risk groups: a cost-utility analysis. Ann Intern Med 2003; 138:176–186.
23. Sampliner RE, Garewal HS, Fennerty MB, Aickin M. Lack of impact of therapy on extent of Barrett's esophagus in 67 patients. Dig Dis Sci 1990;35:93–96.
24. Peters FT, Ganesh S, Kuipers EJ, Sluiter WJ, Klinkenberg-Knol EC, Lamers CB, Kleibeuker JH. Endoscopic regression of Barrett's oesophagus during omeprazole treatment: a randomised double-blind study. Gut 1999;45:489–494.
25. Wilkinson SP, Biddlestone L, Gore S, Shepherd NA. Regression of columnar-lined (Barrett's) oesophagus with omeprazole 40 mg daily: results of 5 years of continuous therapy. Aliment Pharmacol Ther 1999;13:1205–1209.

Received May 18, 2006. Accepted July 28, 2006.

Address requests for reprints to: Prateek Sharma, MD, University of Kansas School of Medicine, Department of Veterans Affairs Medical Center, 4801 East Linwood Boulevard, Kansas City, Missouri 64128-2295. e-mail: psharma@kumc.edu; fax: (816) 922 4692.

Supported by an unrestricted educational grant from AstraZeneca.

The authors thank R. Ackroyd, A. Barkun, H. W. Boyce, A. Clarke, G. Costamagna, A. Edebo, R. Enns, G. Falk, P. Fockens, J. Hatlebakk, R. Holloway, G. Holtmann, P. Katelaris, A. Kruse, N. Marcon, G. May, S. Nandurkar, H. Neuhaus, M-A. Ortner, G. Overholt, U. Peitz, J. Regula, J-F. Rey, R. Sampliner, M. Shoeman, S. Spechler, N. B. Vakil, J. L. van Laethem, and K. Wang for their valuable contributions to this study and Mary Mackison and Francesca Semiera for editorial assistance.