



CLINICAL RESEARCH

Web-based system for training and dissemination of a magnification chromoendoscopy classification

Mário Jorge Dinis-Ribeiro, Ricardo Cruz Correia, Cristina Santos, Sónia Fernandes, Ernesto Palhares, Rui Almeida Silva, Pedro Amaro, Miguel Areia, Altamiro Costa-Pereira, Luis Moreira-Dias

Mário Jorge Dinis-Ribeiro, Rui Almeida Silva, Luis Moreira-Dias, Gastroenterology Department, Portuguese Oncology Institute, Porto 4200-072, Portugal

Mário Jorge Dinis-Ribeiro, Ricardo Cruz Correia, Cristina Santos, Sónia Fernandes, Ernesto Palhares, Altamiro Costa-Pereira, Center for Research in Information Systems and Health Technologies (CINTESIS)/Department of Biostatistics and Medical Informatics, Porto Faculty of Medicine, Porto 4200-319, Portugal

Pedro Amaro, Miguel Areia, Gastroenterology Department, Coimbra University Hospital, Coimbra 3000-075, Portugal

Author contributions: Dinis-Ribeiro M was responsible for the design, implementation, data analysis and writing of the manuscript; Correia R, Fernandes S and Palhares E developed and implemented the web-system, its improvement and critically reviewed the manuscript; Santos C and Costa-Pereira A participated in the design of study and data analysis; Silva RA, Amaro P and Areia M participated in the study and critically reviewed the manuscript for discussion; Moreira-Dias L participated in the design, implementation and discussion of results.

Supported by Sociedade Portuguesa de Endoscopia Digestiva (Research Grant 2002) and the European Society for Gastrointestinal Endoscopy

Correspondence to: Mário Jorge Dinis-Ribeiro, MD, PhD, Gastroenterology Department, Portuguese Oncology Institute, Rua Dr. Bernardino de Almeida, Porto 4200-072, Portugal. mario@med.up.pt

Telephone: +351-22-5084000-3348 Fax: +351-22-5084055

Received: July 28, 2008 Revised: September 9, 2008

Accepted: September 16, 2008

Published online: December 14, 2008

intra- [Cohen's kappa (κ) = 0.79-1.00 to 0.89-1.00] and inter-observer agreement increased from 1st (moderate) to 6th observation (κ = 0.94). Also, agreement with reference increased in the last observations (0.90, 1.00 and 1.00, for observers A, B and C, respectively). Validity of 100% was obtained by all observers at their 4th observation. When a 4th (sub)group was considered, inter-observer agreement was almost perfect (κ = 0.92) at 6th observation. The relation with reference clearly improved into κ (0.93-1.00) and sensitivity (75%-100%) at their 6th observations.

CONCLUSION: This MC classification seems to be easily explainable and learnable as shown by excellent intra- and inter-observer agreement, and improved agreement with reference. A web system such as the one used in this study may be useful for endoscopic or other image based diagnostic procedures with respect to definition, education and dissemination.

© 2008 The WJG Press. All rights reserved.

Key words: Magnification; Chromoendoscopy; Reproducibility; Learning curve

Peer reviewer: Dr. William Dickey, Altnagelvin Hospital, Londonderry, BT47 6SB, Northern Ireland, United Kingdom

Dinis-Ribeiro M, Correia R, Santos C, Fernandes S, Palhares E, Silva RA, Amaro P, Areia M, Costa-Pereira A, Moreira-Dias L. Web-based system for training and dissemination of a magnification chromoendoscopy classification. *World J Gastroenterol* 2008; 14(46): 7086-7092 Available from: URL: <http://www.wjgnet.com/1007-9327/14/7086.asp> DOI: <http://dx.doi.org/10.3748/wjg.14.7086>

Abstract

AIM: To evaluate the use of web-based technologies to assess the learning curve and reassess reproducibility of a simplified version of a classification for gastric magnification chromoendoscopy (MC).

METHODS: As part of a multicenter trial, a hybrid approach was taken using a CD-ROM, with 20 films of MC lasting 5 s each and an "autorun" file triggering a local HTML frameset referenced to a remote questionnaire through an Internet connection. Three endoscopists were asked to prospectively and independently classify 10 of these films randomly selected with at least 3 d apart. The answers were centrally stored and returned to participants together with adequate feedback with the right answer.

RESULTS: For classification in 3 groups, both

INTRODUCTION

The dissemination and teaching of image based medical technologies depend on adequate training. Mostly, medical doctors perform specific training by visiting experts. New information technologies, namely those based on the internet, may circumvent such difficulties at least at early phases of training.

Gastric cancer is the second most lethal cancer in the World. Early stages at diagnosis are related to better prognosis^[1]. Minute flat non-invasive neoplastic lesions

(dysplasia)^[2] may be found during screening programs (in Japan) or during the follow-up of patients with atrophic chronic gastritis (ACG) or intestinal metaplasia (IM)-the milieu where neoplastic changes develop^[3-6]. However, for patients with lesions such as ACG or IM there is not a definite proposal for their management. The difficulty in proposing a guide of practice for the management of patients with atrophic chronic gastritis and intestinal metaplasia may be related to the fact that conventional endoscopy used in most studies shows a low reproducibility and poor relation with histology at diagnosing these diffuse mucosal changes and minute lesions of cancer^[7,8]. For the last ten years, several studies considered magnification and high-resolution endoscopy in conjunction with chromoendoscopy for the diagnosis of precancerous^[9-14] and neoplastic lesions^[15-22], in the gastrointestinal tract. However, mostly authors focused on the validity assessment and rarely were reliability or the learning of each group description defined^[16,17]. Furthermore, several classifications were defined and the need for standardization stressed^[23]. In fact, aimed at improving the evaluation of patients with precancerous gastric conditions, our own group described a classification for the diagnosis of intestinal metaplasia and minute dysplastic lesions using magnification chromoendoscopy with methylene blue^[17].

As part of a multicenter trial, the training of endoscopists and teaching of this classification was planned using a web-based system. This manuscript reports the feasibility of such a system for the learning and dissemination of endoscopic classifications.

MATERIALS AND METHODS

Study design

Three endoscopists (A, B and C), independently and blinded to other endoscopists' answers, were asked to prospectively classify 20 endoscopic videos of magnification chromoendoscopy using a web-based learning system, a hybrid system composed of a CD-ROM and a dynamic website connected to a database, aiming at classifying each video 6 times (1st to 6th time).

Endoscopic videos selection

Endoscopic videos were selected according to a modified version of a magnification chromoendoscopy pattern classification of gastric mucosa^[17]. For video selection the records of magnification chromoendoscopy with methylene blue (1%) using an Olympus Q240Z magnification endoscope (Olympus Corp., Tokyo, Japan), performed in a cohort of patients under follow-up at our institution, were used^[6]. Videos were recorded using a S-Video interface with a digital DVCAM Sony Recorder (DSR-20MDP, Sony, Tokyo, Japan). Endoscopic patterns were obtained using the maximum magnification power possible with this endoscope, defined according to differences in color and homogeneity: Group I definition was when the mucosa showed a regular mucosal pattern and no change in color after staining with methylene blue; Group II if the mucosa presented a regular pattern and was stained

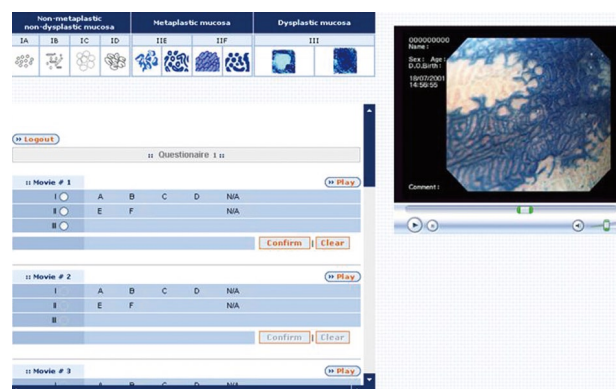


Figure 1 Graphical user interface for movie classification. The right frame is used to play a video clip loaded from the distributed CD; the left frame shows the patterns at the top and a form to retrieve user classifications.

in blue. Subgroup II E, for areas of mucosa with blue irregular marks [initially (Dinis Ribeiro GIE 2003) called II A] or blue round and tubular pits (II B); and subgroup II F when blue villi (formerly II C) or blue small pits (II D) were described in the observed mucosa; Group III was the definition if neither a clear pattern nor a change in color (heterogeneous staining) were noticeable.

Web-based learning system

The expected download time for each 5 s film (Windows Media Player video clips with 36 Mbytes) of about 120 min (at a 56 Kbits/s connection) and the user's physical location were instrumental in choosing the hybrid system architecture^[24,25]. Using an Internet connection, a CD-ROM including all 20 selected endoscopic movies and an "autorun" file used to trigger a local HTML frameset (with two frames) referenced to the remote questionnaire on the classification of each film were developed for this project (on the left)^[26]. The right frame was used to play the films stored on the CD-ROM (Figure 1). At the top, a schematic representation of each pattern was always visible.

Each endoscopist was asked to classify 10 videos randomly selected from the 20 videos included in the CD, with a minimum interval of 3 d. Before classifying each video, the user could run the film as many times as necessary before the decision was taken. After deciding, the user had to lock his answer in order to advance to the next question, not allowing subsequent videos to influence previous responses. After each questionnaire, ie for each 10 film sets classified, their answer or classification together with the proposed answer (to be used as reference, see below) was returned to participants. By this time, all videos included in that set could again be seen.

The HTML questionnaire was stored on an Oracle database using a PHP script. The web-server is run on RedHat Linux 7.2 (Enigma), Apache 1.3, PHP 4.0 compiled with GD Graphics Library 1.8 and Oracle 8 DBMS. Answers were centrally stored on an Oracle database using a Hypertext Pre-Processor (PHP) script.

Endoscopists

Endoscopists (M.A., P.A., R.S.) were invited to participate as they belonged to two different centers (POI

Table 1 Reproducibility for the classification in groups (I vs II vs III) and in subgroups (I vs II E vs II F vs III) according to number of times observers (A, B and C) classified the films of magnification chromoendoscopy

Number of classification (nth)	Inter-observer agreement (95% CI)	
	Proportion of agreement	Weighted kappa
Classification in groups		
1st observation	0.60 (0.36-0.81)	0.52 (0.26-0.75)
2nd observation	0.85 (0.62-0.97)	0.73 (0.53-0.87)
3rd observation	0.85 (0.62-0.97)	0.71 (0.51-0.86)
4th observation	0.95 (0.75-1.00)	0.97 (0.94-0.99)
5th observation	0.80 (0.56-0.94)	0.71 (0.50-0.86)
6th observation	0.90 (0.68-0.99)	0.94 (0.89-0.98)
Classification in subgroups		
1st observation	0.45 (0.23-0.69)	0.49 (0.23-0.73)
2nd observation	0.55 (0.31-0.77)	0.69 (0.48-0.85)
3rd observation	0.45 (0.23-0.69)	0.66 (0.42-0.83)
4th observation	0.70 (0.46-0.88)	0.90 (0.81-0.96)
5th observation	0.60 (0.36-0.81)	0.66 (0.43-0.83)
6th observation	0.75 (0.51-0.91)	0.92 (0.84-0.96)

in Porto, and CUH in Coimbra) inclined to implement this technology, but with no previous experience of it or without previous participation in the development of the classification.

Statistical analysis

For each image, proposed classification (Group I, Group II, Subgroup II E or II F, or Group III) was considered as if another observer would have classified it, and, also, as a reference classification or gold standard. This allowed us to consider both agreement and validity measures, respectively, in the evaluation of reproducibility and learning curve.

Inter-observer agreement and agreement with the reference agreement were estimated using different measures of agreement^[27], simple proportions of agreement (Pa) and proportions of specific agreement, and quadratic weighted Cohen's kappa coefficient (Kc) (estimated by intra-class correlation coefficient)^[28-33]. The confidence intervals for proportions of agreements were estimated with binomial distribution^[33]. Strength of agreement was considered as follows: 0.01-0.2 slight, 0.21-0.4 fair, 0.41-0.6 moderate, 0.61-0.8 substantial, 0.81-1 almost perfect^[34]. No bias was observed [McNemar test ($P = 1.0$); bias index (0.117, $P = 0.2891$)]^[35,36].

Estimates of sensitivity (Se), specificity (Sp) and validity were also calculated comparing the classification for each film against the proposed classified as reference. For classification in groups, true positives were defined if the observer correctly classified each film as group III. For classification in subgroups, diagnostic positivity was considered in cases of Subgroups II F and III. These options and the decision to use weighted kappa coefficient were based upon the relation of these patterns with both dysplasia and incomplete intestinal metaplasia, named as being high-risk lesions for adenocarcinoma^[4-6].

The learning curve was defined by visual analysis of a plot of both validity and agreement measures. Statistical estimates were performed with r-project v2.1.1, SPSS® and MedCalc®.

RESULTS

Reproducibility

Both classification in groups (I vs II vs III) and subgroups (I vs II E vs II F vs III) showed substantial to excellent inter-observer agreement. In fact, at 6th observation, proportion of agreement is 0.90 and 0.75 respectively and weighted kappa is 0.94 and 0.94, respectively (Table 1).

As far as intra-observer agreement is concerned this was substantial in all observers, initially from first to second observation ($\kappa = 0.79$ to 1.00), and excellent from 5th to 6th observation (0.89 to 1) considering the classification in Groups (I vs II vs III); and for classification in subgroups (I vs II E vs II F vs III), 0.74-0.85 to 0.75-1.00.

Specific proportions of agreement were also very high varying from 0.43, 0.79 and 0.82 (for groups III, II and I, respectively) to 0.96, 0.92 and 0.92 (III, II and I) at last classification. Concerning specific proportions of agreement, only a slight increase was observed from 0.50 and 0.50 (II E and II F) to 0.64 and 0.60 (at last observation).

Learning curve

An increase was observed in both proportions of agreement and kappa values, as far as agreement with original classification was concerned, from moderate to substantial/excellent in all observers (Table 2). Also inter-observer agreement varied from Kc = 0.52 or 0.49, respectively for groups and subgroups classification, from 1st to 6th classification (Figure 2). Excellent agreement was obtained by the 4th time for all observers irrespective of institution or time between classifications, by the time they had evaluated 80 videos.

Also, concerning validity measures, paired sensitivity and specificity of 100% were achieved at 4th classification for all observers, at 4th time for classification in groups and at 6th for classification in subgroups by observer A. Observers B and C achieved a validity of 0.85 and 0.90 at their 6th classification (Figure 2).

DISCUSSION

The concept of 'learning by doing' in invasive procedures such as endoscopy, even though current and acceptable, may be affected by the continuous research in this field leading to new endoscopes and gastrointestinal mucosal description availability.

In a preliminary form, we have described the feasibility of a hybrid approach of Internet and CD-ROM/DVD technology as a web-based education system^[37]. Such desktop virtual reality systems^[38] were described in several fields of knowledge^[24,25] as recently in endoscopy by de Lange^[39].

According to our study, the classification proposed is both easily explainable and learnable. The simplicity of this classification, the fact that it includes in the instrument description the phenomenon itself (i.e. intestinal metaplasia) and the feedback given to each observer at the end of a single classification^[40,41] may

Table 2 Agreement with reference and validity measures for the classification in groups (I *vs* II *vs* III) and in subgroups (I *vs* II *vs* III) according to number of times observers (A, B and C) classified the films of magnification chromoendoscopy (95% CI)

	Classification in groups I <i>vs</i> II <i>vs</i> III					Classification in Subgroups I <i>vs</i> II <i>vs</i> III				
	Pa	wK	Se	Sp	V	Pa	wK	Se	Sp	V
Observer A										
1st observation	0.90 (0.68-0.99)	0.66 (0.32-0.85)	0.75 (0.56-0.94)	0.94 (0.83-1.00)	0.90 (0.77-1.00)	0.75 (0.51-0.91)	0.63 (0.29-0.84)	0.75 (0.56-0.94)	0.75 (0.56-0.94)	0.75 (0.56-0.94)
2nd observation	0.95 (0.75-1.00)	0.84 (0.64-0.93)	1.00	0.94 (0.83-1.00)	0.95 (0.85-1.00)	0.70 (0.46-0.88)	0.76 (0.50-0.90)	0.75 (0.56-0.94)	0.67 (0.46-0.88)	0.70 (0.50-0.90)
3rd observation	0.90 (0.68-0.99)	0.79 (0.54-0.91)	1.00	0.94 (0.83-1.00)	0.95 (0.85-1.00)	0.80 (0.56-0.94)	0.79 (0.55-0.91)	1.00	0.75 (0.56-0.94)	0.85 (0.69-1.00)
4th observation	1.00 (0.86-1.00)	1.00	1.00	1.00	1.00	0.90 (0.68-0.99)	0.97 (0.92-0.99)	1.00	0.83 (0.67-1.00)	0.90 (0.85-1.00)
5th observation	0.95 (0.75-1.00)	0.83 (0.63-0.93)	1.00	0.94 (0.83-1.00)	0.95 (0.85-1.00)	0.95 (0.75-1.00)	0.85 (0.66-0.94)	1.00	0.92 (0.79-1.00)	0.95 (0.85-1.00)
6th observation	1.00 (0.86-1.00)	1.00	1.00	1.00	1.00	1.00 (0.86-1.00)	1.00	1.00	1.00	1.00
Observer B										
1st observation	0.90 (0.51-0.91)	0.80 (0.57-0.92)	0.75 (0.56-0.94)	1.00	0.95 (0.85-1.00)	0.65 (0.41-0.85)	0.71 (0.42-0.88)	0.50 (0.28-0.72)	0.83 (0.67-1.00)	0.70 (0.50-0.90)
2nd observation	0.90 (0.68-0.99)	0.77 (0.52-0.90)	0.75 (0.56-0.94)	1.00	0.95 (0.85-1.00)	0.80 (0.56-0.94)	0.78 (0.52-0.90)	0.75 (0.56-0.94)	0.92 (0.79-1.00)	0.85 (0.69-1.00)
3rd observation	0.90 (0.68-0.99)	0.77 (0.52-0.90)	0.75 (0.56-0.94)	1.00	0.95 (0.85-1.00)	0.70 (0.46-0.88)	0.69 (0.37-0.86)	0.63 (0.41-0.84)	0.75 (0.56-0.94)	0.70 (0.50-0.90)
4th observation	0.95 (0.75-1.00)	0.96 (0.90-0.98)	1.00	1.00	1.00	0.65 (0.41-0.85)	0.82 (0.60-0.92)	0.5 (0.28-0.72)	0.75 (0.56-0.94)	0.65 (0.44-0.86)
5th observation	0.90 (0.68-0.99)	0.77 (0.52-0.90)	0.75 (0.56-0.94)	1.00	0.95 (0.85-1.00)	0.65 (0.41-0.85)	0.73 (0.45-0.88)	0.63 (0.41-0.84)	0.92 (0.79-1.00)	0.80 (0.62-0.98)
6th observation	1.00 (0.86-1.00)	1.00	0.75 (0.56-0.94)	1.00	1.00	0.85 (0.62-0.97)	0.95 (0.87-0.98)	0.75 (0.56-0.94)	0.92 (0.79-1.00)	0.85 (0.69-1.00)
Observer C										
1st observation	0.75 (0.51-0.91)	0.80 (0.57-0.92)	0.75 (0.56-0.94)	0.81 (0.64-0.99)	0.80 (0.62-0.98)	0.60 (0.36-0.81)	0.82 (0.60-0.92)	0.75 (0.56-0.94)	0.83 (0.67-1.00)	0.80 (0.62-0.98)
2nd observation	0.95 (0.75-1.00)	0.96 (0.89-1.00)	1.00	1.00	1.00	0.60 (0.36-0.81)	0.85 (0.67-0.94)	0.50 (0.28-0.72)	0.75 (0.56-0.94)	0.65 (0.44-0.86)
3rd observation	0.85 (0.62-0.97)	0.72 (0.42-0.88)	1.00	0.84 (0.67-1.00)	0.85 (0.69-1.00)	0.50 (0.27-0.73)	0.65 (0.31-0.84)	0.38 (0.16-0.59)	0.75 (0.56-0.94)	0.60 (0.38-0.82)
4th observation	0.95 (0.75-1.00)	0.96 (0.89-0.98)	1.00	1.00	1.00	0.70 (0.46-0.88)	0.89 (0.74-0.95)	0.5 (0.28-0.72)	0.92 (0.79-1.00)	0.75 (0.56-0.94)
5th observation	0.85 (0.62-0.97)	0.74 (0.47-0.89)	0.75 (0.56-0.94)	0.94 (0.83-1.00)	0.90 (0.77-1.00)	0.65 (0.41-0.85)	0.73 (0.45-0.88)	0.63 (0.41-0.84)	0.83 (0.67-1.00)	0.75 (0.56-0.94)
6th observation	0.90 (0.68-0.99)	0.92 (0.81-0.97)	1.00	0.94 (0.83-1.00)	0.95 (0.85-1.00)	0.80 (0.56-0.94)	0.93 (0.84-0.97)	0.88 (0.73-1.00)	0.92 (0.79-1.00)	0.90 (0.77-1.00)

Pa: Proportion of agreement; wK: Weighted kappa; Se: Sensitivity; V: Validity.

justify the excellent results.

Learning curves for most procedures concern efficacy and time to achievement of such efficacy. For example, in surgical procedures how fast trainees achieve the ability to get surgery adequately performed without complications^[42,43]. Also in endoscopy some reports use colonoscopy models^[44] and endoscopic ultrasound fine needle aspiration^[45] with similar methodology.

A single report exists on the learning curve for the diagnostic performance of endoscopy. Besides simplification of any visual categorization, Tung and Tagashi defined the need of a steep learning curve for magnifying colonoscopy. They used for that evaluation the evolution of validity measures, that is sensitivity, specificity and global accuracy. However, there is no one particular statistical procedure for learning assessment, to be named in diagnostic technologies outcomes as a measure of reality^[46].

Diagnostic procedures are aimed at being both valid, ie to measure what they are supposed to. However,

even though most studies concern validity assessment, reliability of a measure should be a condition to be verified before any other quality feature. For dichotomic, nominal or ordinal variables, proportion of agreement (Pa) or Kappa statistics may be used. Pa is easily acceptable and interpretable. However, it is not corrected to the amount of agreement that was expected by chance (Pe)^[47]. With the aim of solving this problem, Cohen developed the named Kappa statistics (Kc); a method which takes into consideration the so-called agreement by chance. In ordinal variables, distance from total agreement may be weighted, either linearly or exponentially^[48]. Therefore, kappa is a global index of agreement and easy to calculate. However, some concern has been raised by others^[48,49]. Cohen's Kappa varies with the distribution of cases for each category, namely as far as the total number of cases or prevalence^[50] and if unbalanced marginal totals is present. Additionally, bias can influence kappa's interpretation^[51]. Therefore, some authors recommend the estimation of both prevalence

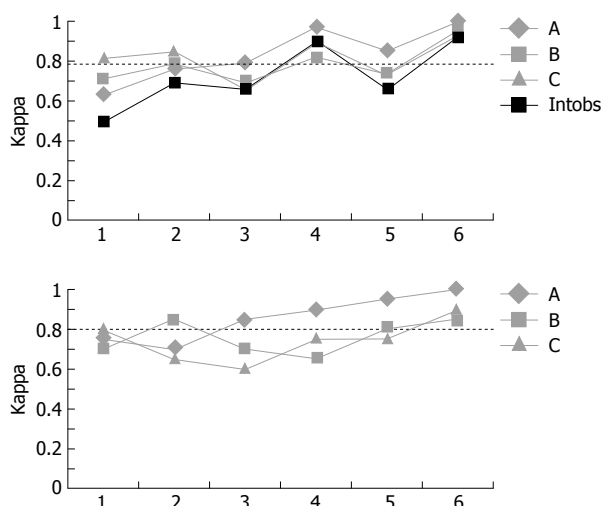


Figure 2 Variation of agreement with reference (Kappa) and inter-observer agreement (Black) (top graphic) and validity (bottom graphic) along sequential observations (1st to 6th) for the classification in subgroups [dashed line marks for 0.80 as the cutoff for almost perfect agreement (top graphic) and validity (bottom graphic)].

and bias adjusted kappa^[35,36] and others advocate the use of either McNemar's test or the bias index (proportion of deviated ratings) to assess bias, following which Cohen's kappa could then be used. It seems reasonable to consider that both agreement measures (proportion of agreement and weighted Cohen's kappa) and validity measures. Thus, the proposed original classification could be considered either as the classification by a different observer and agreement with it by each observer would be evaluated for reliability or it may be considered as reference and common measures of validity would be used similarly to the paper of Tung and Tagashi.

In the present study, although 20 selected non-consecutive films were assessed, no observer bias was noticeable and the fact that all categories of classification groups were included for evaluation, may allow us, even though cautiously, to consider our classification of gastric mucosa both reliable and easily learnable.

The follow-up of patients with atrophic chronic gastritis and intestinal metaplasia^[52,53] may lead to early diagnosis of gastric neoplastic lesions and improvement of patients' prognosis. Following the non-existence of distinctive symptoms^[54-56], most authors based their studies (mostly) on morphologic evaluation through endoscopically performed multiple biopsies, because of the patchy characteristics of atrophic chronic gastritis and intestinalization of gastric mucosa^[57-59]. However, with the exception of atrophic vascularization, most studies found that for conventional endoscopy, descriptions of 'gastritis' showed suboptimal validity^[60-63] and unsatisfactory reliability^[7,8,63].

New endoscopic methods are expected to optimise both the identification, in a (more) reproducible and valid measure for such lesions-'optic biopsy'. An increasing number of expert opinion texts, reviews and studies report the use of magnification chromoendoscopy through the gastrointestinal tract.

As far as colorectal lesions are concerned, in 1996 Kudo *et al*^[15] defined a 7 patterns classification (type I, II, IIIs, III L, IV, Vn, Vi) that showed consistently good sensitivity but highly heterogeneous results in its specificity^[15,22]. Eight years have past and recently reproducibility was demonstrated^[16], in an altered simplified three patterns classification with management consequences or prognosis implications: I and II as non-neoplastic; III L and IV as neoplastic; and III s and V as neoplastic possibly invasive.

However, in upper gastrointestinal tract, both for Barrett's mucosa and stomach mucosa, diverse classifications have been published and the need for their standardization stressed^[23].

Endo *et al*^[10] and Yagi *et al*^[21] using methylene blue, Guelrud *et al*^[12,14] and Toyoda *et al*^[64] using acetic acid, and Sharma *et al*^[13] using indigo carmine, described features of intestinal metaplasia and Sharma also reported endoscopic dysplasia. Good validity results were published by all authors, but Meining *et al*^[11] showed a low inter-observer agreement, both for Endo and for Guelrud classifications (Cohen's kappa of 0.017 and 0.162).

In the stomach, our own group described the use of magnification chromoendoscopy with methylene blue for the diagnosis of intestinal metaplasia and gastric epithelial dysplasia in 2003. We subsequently found that a substantial agreement was observed on the classification of endoscopic images into groups (I, II, or III), both for intra-observer (Pa = 0.91, Kc = 0.86) and for inter-observer agreement (Pa = 0.84, Kc = 0.74)^[17]. Hereby, the stomach size and the presence of inflammation were considered limitations for chromoendoscopy and particularly for magnification^[65].

However, concurrent results by others working in the field of gastric mucosa were consistent with ours. Recently, Yagi *et al*^[66] described aspects for normal antral mucosa and for gastritis with H pylori similar to our group I. Also Yang types A through D^[67] and Kim types 1 through 3^[18] may be compared with our group classification as Group I. Furthermore, Kim's type 4 and Yang types D and E are very similar to Subgroups II E and II F. Tajiri *et al*^[19,20] stressed that this procedure may have marked impact in the diagnosis of minute neoplastic flat 'gastritis-like' lesions and they described very similar features to our own research group's Group III or pattern-less.

This means that, as with Kudo's classification in the colon, the existence of a unique and standardized classification for magnification chromoendoscopy (both in Barrett's and in the stomach) may have contributed to the dissemination of this technique and further use in even newer technologies.

In conclusion, a modified version of our classification for gastric mucosa diffuse changes and minute dysplastic lesions seems to be reliable and easily learnable. The web-based system hereby developed can be used for new diagnostic technology teaching and dissemination and for assessing the similarity between our own and other classifications, with the aim of the achievement of consensus.

ACKNOWLEDGMENTS

This project and preliminary results were presented in a Poster Session at the 8th Annual World Congress of the Internet and Medicine (MedNet), Geneva, Switzerland (2003), published as an abstract in the *Int J Health Care Engineering* 2003; 11: 371-372, in a Poster Session at Digestive Diseases Week in Chicago in May 2005 and published in *Gastrointestinal Endoscopy* as an abstract.

COMMENTS

Background

Dissemination and teaching of image-based medical technologies depend on adequate training. Mostly, medical doctors perform specific training by visiting experts. New information technologies, namely those based on the internet, may circumvent such difficulties at least at early phases of training.

Research frontiers

Concerning digestive endoscopy, several studies addressed and derived classifications for endoscopic descriptions of precancerous and neoplastic lesions in the gastrointestinal tract. Web-based systems could also be used in this setting for dissemination and training.

Innovations and breakthroughs

As part of a multicenter trial, the training of endoscopists and teaching of this classification were planned using a web-based system. This manuscript reports the feasibility of such a system for the learning and dissemination of endoscopic classifications.

Applications

Similar systems could be used in other areas of medical technologies based on image. Furthermore, similar methodologies could even go further by gathering clinical data, other technologies could be used in medical decision analysis.

Peer review

This manuscript specifically addresses a new evaluation of the reliability of a classification for magnification chromoendoscopy. It describes the feasibility of a web-based system to be used as part of endoscopists' training in learning new technologies.

REFERENCES

- 1 Levi F, La Vecchia C, Lucchini F, Negri E. Cancer mortality in Europe, 1990-1992. *Eur J Cancer Prev* 1995; **4**: 389-417
- 2 Schlemper RJ, Riddell RH, Kato Y, Borchard F, Cooper HS, Dawsey SM, Dixon MF, Fenoglio-Preiser CM, Flejou JF, Geboes K, Hattori T, Hirota T, Itabashi M, Iwafuchi M, Iwashita A, Kim YI, Kirchner T, Klimpfinger M, Koike M, Lauwers GY, Lewin KJ, Oberhuber G, Offner F, Price AB, Rubio CA, Shimizu M, Shimoda T, Sipponen P, Solcia E, Stolte M, Watanabe H, Yamabe H. The Vienna classification of gastrointestinal epithelial neoplasia. *Gut* 2000; **47**: 251-255
- 3 Correa P. Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Res* 1992; **52**: 6735-6740
- 4 Carneiro F, Machado JC, David L, Reis C, Nogueira AM, Sobrinho-Simoes M. Current thoughts on the histopathogenesis of gastric cancer. *Eur J Cancer Prev* 2001; **10**: 101-102
- 5 Kapadia CR. Gastric atrophy, metaplasia, and dysplasia: a clinical perspective. *J Clin Gastroenterol* 2003; **36**: S29-S36; discussion S61-S62
- 6 Dinis-Ribeiro M, Lopes C, da Costa-Pereira A, Guilherme M, Barbosa J, Lomba-Viana H, Silva R, Moreira-Dias L. A follow up model for patients with atrophic chronic gastritis and intestinal metaplasia. *J Clin Pathol* 2004; **57**: 177-182
- 7 Laine L, Cohen H, Sloane R, Marin-Sorensen M, Weinstein WM. Interobserver agreement and predictive value of endoscopic findings for H. pylori and gastritis in normal volunteers. *Gastrointest Endosc* 1995; **42**: 420-423
- 8 Belair PA, Metz DC, Faigel DO, Furth EE. Receiver operator characteristic analysis of endoscopy as a test for gastritis. *Dig Dis Sci* 1997; **42**: 2227-2233
- 9 Bruno MJ. Magnification endoscopy, high resolution endoscopy, and chromoscopy; towards a better optical diagnosis. *Gut* 2003; **52** Suppl 4: iv7-ii11
- 10 Endo T, Awakawa T, Takahashi H, Arimura Y, Itoh F, Yamashita K, Sasaki S, Yamamoto H, Tang X, Imai K. Classification of Barrett's epithelium by magnifying endoscopy. *Gastrointest Endosc* 2002; **55**: 641-647
- 11 Meining A, Rosch T, Kiesslich R, Muders M, Sax F, Heldwein W. Inter- and intra-observer variability of magnification chromoendoscopy for detecting specialized intestinal metaplasia at the gastroesophageal junction. *Endoscopy* 2004; **36**: 160-164
- 12 Guelrud M, Herrera I, Essenfled H, Castro J. Enhanced magnification endoscopy: a new technique to identify specialized intestinal metaplasia in Barrett's esophagus. *Gastrointest Endosc* 2001; **53**: 559-565
- 13 Sharma P, Weston AP, Topalovski M, Cherian R, Bhattacharyya A, Sampliner RE. Magnification chromoendoscopy for the detection of intestinal metaplasia and dysplasia in Barrett's oesophagus. *Gut* 2003; **52**: 24-27
- 14 Guelrud M, Herrera I, Essenfled H, Castro J, Antonioli DA. Intestinal metaplasia of the gastric cardia: A prospective study with enhanced magnification endoscopy. *Am J Gastroenterol* 2002; **97**: 584-589
- 15 Kudo S, Tamura S, Nakajima T, Yamano H, Kusaka H, Watanabe H. Diagnosis of colorectal tumorous lesions by magnifying endoscopy. *Gastrointest Endosc* 1996; **44**: 8-14
- 16 Huang Q, Fukami N, Kashida H, Takeuchi T, Kogure E, Kurahashi T, Stahl E, Kudo Y, Kimata H, Kudo SE. Interobserver and intra-observer consistency in the endoscopic assessment of colonic pit patterns. *Gastrointest Endosc* 2004; **60**: 520-526
- 17 Dinis-Ribeiro M, da Costa-Pereira A, Lopes C, Lara-Santos L, Guilherme M, Moreira-Dias L, Lomba-Viana H, Ribeiro A, Santos C, Soares J, Mesquita N, Silva R, Lomba-Viana R. Magnification chromoendoscopy for the diagnosis of gastric intestinal metaplasia and dysplasia. *Gastrointest Endosc* 2003; **57**: 498-504
- 18 Kim S, Harum K, Ito M, Tanaka S, Yoshihara M, Chayama K. Magnifying gastroendoscopy for diagnosis of histologic gastritis in the gastric antrum. *Dig Liver Dis* 2004; **36**: 286-291
- 19 Tajiri H, Doi T, Endo H, Nishina T, Terao T, Hyodo I, Matsuda K, Yagi K. Routine endoscopy using a magnifying endoscope for gastric cancer diagnosis. *Endoscopy* 2002; **34**: 772-777
- 20 Tajiri H, Ohtsu A, Boku N, Muto M, Chin K, Matsumoto S, Yoshida S. Routine endoscopy using electronic endoscopes for gastric cancer diagnosis: retrospective study of inconsistencies between endoscopic and biopsy diagnoses. *Cancer Detect Prev* 2001; **25**: 166-173
- 21 Yagi K, Nakamura A, Sekine A. Accuracy of magnifying endoscopy with methylene blue in the diagnosis of specialized intestinal metaplasia and short-segment Barrett's esophagus in Japanese patients without Helicobacter pylori infection. *Gastrointest Endosc* 2003; **58**: 189-195
- 22 Tung SY, Wu CS, Su MY. Magnifying colonoscopy in differentiating neoplastic from nonneoplastic colorectal lesions. *Am J Gastroenterol* 2001; **96**: 2628-2632
- 23 Sharma P. Magnification endoscopy. *Gastrointest Endosc* 2005; **61**: 435-443
- 24 Bacro T, Gilbertson B, Coultas J. Web-delivery of anatomy video clips using a CD-ROM. *Anat Rec* 2000; **261**: 78-82
- 25 Mattheos N, Nattestad A, Attstrom R. Local CD-ROM in interaction with HTML documents over the Internet. *Eur J Dent Educ* 2000; **4**: 124-127
- 26 Cruz-Correia R, Dinis-Ribeiro M, Fernandes S, Oliveira-Palhares E, Martins C, Costa-Pereira A. ALGA a Web-based gastrointestinal endoscopy learning curve evaluation

- system. *Technol Health Care* 2003; **11**: 371-372
- 27 **Costa Santos C**, Costa Pereira A, Bernardes J. Agreement studies in obstetrics and gynaecology: inappropriateness, controversies and consequences. *BJOG* 2005; **112**: 667-669
 - 28 **Cohen J**. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968; **70**: 213-220
 - 29 **Uebersax JS**. A generalized kappa coefficient. *Educ Psychol Meas* 1982; **42**: 181-183
 - 30 **Haley SM**, Osberg JS. Kappa coefficient calculation using multiple ratings per subject: a special communication. *Phys Ther* 1989; **69**: 970-974
 - 31 **Fleiss JL**, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973; **33**: 613-619
 - 32 **Markus H**, Bland JM, Rose G, Sitzer M, Siebler M. How good is intercenter agreement in the identification of embolic signals in carotid artery disease? *Stroke* 1996; **27**: 1249-1252
 - 33 **Clopper CJ**, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934; **26**: 404-413
 - 34 **Landis JR**, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159-174
 - 35 **Ludbrook J**. Statistical techniques for comparing measurers and methods of measurement: a critical review. *Clin Exp Pharmacol Physiol* 2002; **29**: 527-536
 - 36 **Ludbrook J**. Detecting systematic bias between two raters. *Clin Exp Pharmacol Physiol* 2004; **31**: 113-115
 - 37 **Dinis-Ribeiro M**, Cruz-Correia R, Santos C, Fernandes S, Tavares C, Palhares E, Silva RA, Amaro P, Areia M, Ponchon T, Costa-Pereira A, Moreira-Dias L. Reproducibility and learning curve for a classification of magnification chromoendoscopy for gastric mucosal lesions-A web based evaluation. *Gastrointest Endosc* 2005; **12**: 15
 - 38 **Vozenilek J**, Huff JS, Reznick M, Gordon JA. See one, do one, teach one: advanced technology in medical education. *Acad Emerg Med* 2004; **11**: 1149-1154
 - 39 **de Lange T**, Svensen AM, Larsen S, Aabakken L. The functionality and reliability of an Internet interface for assessments of endoscopic still images and video clips: distributed research in gastroenterology. *Gastrointest Endosc* 2006; **63**: 445-452
 - 40 **Mahmood T**, Darzi A. The learning curve for a colonoscopy simulator in the absence of any feedback: no feedback, no learning. *Surg Endosc* 2004; **18**: 1224-1230
 - 41 **Rosser JC Jr**, Gabriel N, Herman B, Murayama M. Telementoring and teleproctoring. *World J Surg* 2001; **25**: 1438-1448
 - 42 **Schlachta CM**, Mamazza J, Seshadri PA, Cadeddu M, Gregoire R, Poulin EC. Defining a learning curve for laparoscopic colorectal resections. *Dis Colon Rectum* 2001; **44**: 217-222
 - 43 **Tekkis PP**, Senagore AJ, Delaney CP, Fazio VW. Evaluation of the learning curve in laparoscopic colorectal surgery: comparison of right-sided and left-sided resections. *Ann Surg* 2005; **242**: 83-91
 - 44 **Eversbusch A**, Grantcharov TP. Learning curves and impact of psychomotor training on performance in simulated colonoscopy: a randomized trial using a virtual reality endoscopy trainer. *Surg Endosc* 2004; **18**: 1514-1518
 - 45 **Eloubeidi MA**, Tamhane A. EUS-guided FNA of solid pancreatic masses: a learning curve with 300 consecutive procedures. *Gastrointest Endosc* 2005; **61**: 700-708
 - 46 **Ramsay CR**, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Statistical assessment of the learning curves of health technologies. *Health Technol Assess* 2001; **5**: 1-79
 - 47 **Bland JM**, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307-310
 - 48 **Chmura Kraemer H**, Periyakoil VS, Noda A. Kappa coefficients in medical research. *Stat Med* 2002; **21**: 2109-2129
 - 49 **Thompson WD**, Walter SD. A reappraisal of the kappa coefficient. *J Clin Epidemiol* 1988; **41**: 949-958
 - 50 **Byrt T**, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993; **46**: 423-429
 - 51 **Khan KS**, Chien PF. Evaluation of a clinical test. I: assessment of reliability. *BJOG* 2001; **108**: 562-567
 - 52 **Genta RM**, Rugge M. Gastric precancerous lesions: heading for an international consensus. *Gut* 1999; **45** Suppl 1: I5-I8
 - 53 **Genta RM**. Review article: Gastric atrophy and atrophic gastritis--nebulous concepts in search of a definition. *Aliment Pharmacol Ther* 1998; **12** Suppl 1: 17-23
 - 54 **Westbrook JL**, McIntosh JH, Duggan JM. Accuracy of provisional diagnoses of dyspepsia in patients undergoing first endoscopy. *Gastrointest Endosc* 2001; **53**: 283-288
 - 55 **Wallace MB**, Durkalski VL, Vaughan J, Palesch YY, Libby ED, Jowell PS, Nickl NJ, Schutz SM, Leung JW, Cotton PB. Age and alarm symptoms do not predict endoscopic findings among patients with dyspepsia: a multicentre database study. *Gut* 2001; **49**: 29-34
 - 56 **Dinis-Ribeiro M**, Lomba-Viana H, Silva R, Fernandes N, Abreu N, Brandao C, Moreira-Dias L, da Costa-Pereira A. Should we exclude individuals from endoscopy based exclusively on the absence of alarm symptoms? *Scand J Gastroenterol* 2004; **39**: 910-911
 - 57 **Rugge M**, Cassaro M, Di Mario F, Leo G, Leandro G, Russo VM, Pennelli G, Farinati F. The long term outcome of gastric non-invasive neoplasia. *Gut* 2003; **52**: 1111-1116
 - 58 **Guarner J**, Herrera-Goepfert R, Mohar A, Sanchez L, Halperin D, Ley C, Parsonnet J. Interobserver variability in application of the revised Sydney classification for gastritis. *Hum Pathol* 1999; **30**: 1431-1434
 - 59 **Dixon MF**, Genta RM, Yardley JH, Correa P. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. *Am J Surg Pathol* 1996; **20**: 1161-1181
 - 60 **Atkins L**, Benedict EB. Correlation of gross gastroscopic findings with gastroscopic biopsy in gastritis. *N Engl J Med* 1956; **254**: 641-644
 - 61 **Heinkel K**. Correlation of gastroscopy, gastric photography and biopsy in diagnosis. *Gastrointest Endosc* 1969; **16**: 81-85
 - 62 **Myren J**, Serck-Hanssen A. The gastroscopic diagnosis of gastritis with particular reference to mucosal reddening and mucus covering. *Scand J Gastroenterol* 1974; **9**: 457-462
 - 63 **Sauerbruch T**, Schreiber MA, Schussler P, Permanetter W. Endoscopy in the diagnosis of gastritis. Diagnostic value of endoscopic criteria in relation to histological diagnosis. *Endoscopy* 1984; **16**: 101-104
 - 64 **Toyoda H**, Rubio C, Befrits R, Hamamoto N, Adachi Y, Jaramillo E. Detection of intestinal metaplasia in distal esophagus and esophagogastric junction by enhanced-magnification endoscopy. *Gastrointest Endosc* 2004; **59**: 15-21
 - 65 **Kiesslich R**, Fritsch J, Holtmann M, Koehler HH, Stoltz M, Kanzler S, Nafe B, Jung M, Galle PR, Neurath MF. Methylene blue-aided chromoendoscopy for the detection of intraepithelial neoplasia and colon cancer in ulcerative colitis. *Gastroenterology* 2003; **124**: 880-888
 - 66 **Yagi K**, Nakamura A, Sekine A. Characteristic endoscopic and magnified endoscopic findings in the normal stomach without *Helicobacter pylori* infection. *J Gastroenterol Hepatol* 2002; **17**: 39-45
 - 67 **Yang JM**, Chen L, Fan YL, Li XH, Yu X, Fang DC. Endoscopic patterns of gastric mucosa and its clinicopathological significance. *World J Gastroenterol* 2003; **9**: 2552-2556

S- Editor Zhong XY L- Editor Logan S E- Editor Lin YP