sciendo Acta Marisiensis - Seria Medica 2023;69(2):104-108 DOI: 10.2478/amma-2023-0015

RESEARCH ARTICLE

Inter-observer variability on the value of endoscopic images for the documentation of upper gastrointestinal endoscopy - our center experience

Ioana Natalia Bernatchi[1*], Septimiu Voidazan[2], Madalina Ioana Petrut[1], Gabriella Gabos[1], Madalin Balasescu[3], Carmen Nicolau[1]

1. Gastroenterology Department, Lotus Image Medical Center, Târgu Mureș, Romania
2. Department of Epidemiology, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania
3. Intensive Care Unit, Mureș County Clinical Hospital, Targu Mures, Romania

Objective: Endoscopy is an essential and invaluable diagnostic tool in the arsenal of every gastroenterologist. ESGE presented additional guidelines for standardized image documentation in upper and lower gastrointestinal endoscopy. Clinical disagreement is a common challenge in most, if not all, fields of medicine. Settling disagreements is important so as to find ways to minimize it. Clinical disagreement in gastroscopy may be demonstrated by studying the observer variability. **Methods**: We retrospectively recruited 120 random patients that underwent conventional upper gastrointestinal endoscopy between 2021-2022 in our Department of Gastroenterology, all of them performed by one endoscopist. As part of the study, all video-endoscopic recordings were stored using one internal server. In order to study interobserver variability, four physicians (endoscopists and gastroenterologist specialists) were invited to complete the questionnaire. **Results**: The interobserver variability in our study ranged from moderate to very good in the assessment of the esophagus, with the highest degree of agreement in response to questions concerning characteristic findings such as normal mucosa, esophagitis Class A Los Angeles, hiatal hernia for the esophagus endoscopic evaluation, benign ulcer niche in gastric antrum, normal gastric corpus mucosa, intestinal metaplasia and angiodysplasia in gastric corpus. The question on atrophic mucosa in the first and second part of the duodenum was the most difficult to agree upon. **Conclusion**: The present study found that the variability between observers in the assessment of images obtained from patients that underwent conventional upper gastrointestinal endoscopy in our center was acceptably good.

Keywords: documentation, endoscopy, gastrointestinal, image, observer variation

Received 31 January 2023 / Accepted 13 March 2023

Introduction

Endoscopy is an essential and priceless diagnostic resource in the arsenal of each gastroenterologist. Considering the fact that the volume as well as the technical complexity of endoscopic procedures has grown over in the past couple of decades, the necessary condition for quality and safety remains crucial [1,2].

Standardization of endoscopic records has been emphasized by the European Society of Gastrointestinal Endoscopy (ESGE) with the development of the minimum standard terminology for digestive endoscopy (MST) [3] subsequently adopted by the World Organization of Gastrointestinal Endoscopy. In addition, ESGE presented additional guidelines for standardized image documentation in upper and lower gastrointestinal endoscopy [4].

Clinical disagreement is a common challenge in most, if not all, fields of medicine. Settling disagreements is important so as to find ways to minimize it. Clinical disagreement in gastroscopy may be demonstrated by studying the observer variability [1].

The aim of the study was to appreciate inter-observer variability in the evaluation of 120 video-endoscopic recordings of conventional upper gastrointestinal endoscopy in our center by four physicians (endoscopists and gastroenterologist specialists).

Methods

We retrospectively recruited 120 random patients that underwent conventional upper gastrointestinal endoscopy between 2021-2022 in our Department of Gastroenterology, all of them performed by one endoscopist. As part of the study, all video-endoscopic recordings were stored using one internal server. In our study we have asked four endoscopists with different level of experience to complete the questionnaire detailed below after assessment of 120 video-recordings.

All subjects were examined using videoendoscope (EG-760R; Fujifilm).

For local anesthesia we commonly used prior to procedure lidocaine spray (4.6mg/dose) and for endoscopic procedures where conscious sedation was performed, midazolam (Midazolam, Aguettant, France) with or without propofol (Propofol MCT/LCT Fresnius 10 mg/ml, Fresnius Kabi) was selected.

All names or dates were removed from the videos. No patient data, characteristics or symptoms were presented.

In order to study inter-observer variability, four physicians (endoscopists and gastroenterologist specialists) with varying endoscopy experience (1 year, 4 and 5 years, 20

* Correspondence to: Ioana Natalia Bernatchi
E-mail: matei_ioana89@yahoo.com

years) were invited to complete the questionnaire. The questions are partially presented in Table I.

Each of the endoscopists evaluated 120 video-recordings and then completed the questionnaire.

Data were collected with a multiple-choice questionnaire containing questions reflecting a simplified version of the minimum standard terminology (MST) for digestive endoscopy, which includes the LA classification for esophagitis. Our interest in this study was the variability between observers in evaluating the images, not if they reached an accurate diagnosis. Therefore, evaluations have not been measured against a "gold standard".

All four endoscopists know and apply in day by day practice the LA classification, they were not given any guidelines on answering the questionnaire.

Ethics

As part of the study, all video-endoscopic recordings were submitted without personal identification, were stored using one internal server and the study was approved by the local Medical Research Ethics Committee.

Statistical analysis

We became interested in this study to evaluate the variability among observers in evaluating images obtained from patients that underwent conventional upper gastrointestinal endoscopy.

The coefficient of agreement for endoscopic diagnosis was evaluated using an inter-rater agreement statistic (K, Kappa) which is calculated with 95% confidence interval [5]. The kappa value was calculated in all of the groups. Agreement, based on the value of kappa, was categorized, as described by Altman, as poor (< 0.20), fair (0.21 - 0.40), moderate (0.41 - 0.60), good (0.61 - 0.80) or very good (0.81 - 1.00) [6]. The precision of kappa was measured by its 95% confidence interval (CI). If the kappa value was greater than 0.40, an acceptable degree of concordance was considered to be present. The analysis was done using SPSS statistical software (SPSS Inc., Chicago, Ill., USA, version 23) for crosstabulation of results and using Excel software (Microsoft Corporation) for measures of kappa value and confidence intervals (CI). Also, nominal variables were described as absolute and relative frequencies (%) and the association between them was analyzed by Pearsons chi square test or Fischer exact test. Associations having P<0.05 were considered to be significant.

Results

Level of agreement is defined as outlined in the Methods section. After assessing video-recordings from the esophagus, 2 question obtained a very good agreement: 0.9668 (CI%:0.81-1.00) on the presence of normal mucosa and the presence of hiatal hernia 0.9711 (CI%:0.81-1.00). Regarding the presence of esophagitis Class D using the Los Angeles classification and ulcer niche, there were no cases identified in the group of 120 randomly selected patients. The interobserver variability in our study ranged from moderate to very good in the assessment of the esophagus, with the highest degree of agreement in response to questions concerning characteristic findings such as normal mucosa, esophagitis Class A Los Angeles, hiatal hernia as can be seen in Table II.

In the assessment of video-recordings from the pyloric antrum, we obtained a very good agreement 0.9682 (CI%:0.81-1.00), in evaluating the presence of the benign ulcer niche. Instead, no cases of angiodysplasia and neoplasia/ malign were identified in the group. The interobserver variability in our study was very good in assessment of the lesions identified in the gastric antrum (Table III).

Table I. Questionnaire as presented to the endoscopists (partial)

Subject	Question	Options
Esophageal images	normal	Yes/No
	mucosal erosion	Yes/No
	Esophagitis according to the LA classification?	None/A/B/C/D
	Hiatal hernia	Yes/No
	Z-line irregularities	Yes/No
	Esophageal varices	Yes/No
	vegetative / proliferative lesion	Yes/No
	ulcer niche	Yes/No

Table II. Results of the evaluation of the oesophagus

Esophagus	Endoscopists				P value*	CI % for K value
	A	B	C	D		
normal	45.0%	41.7%	40.8%	45.0%	0.88	0.95 to 0.97
Esophagitis -Los Angeles class- A	19.2%	22.5%	25.8%	19.2%	0.54	0.91 to 0.95
Esophagitis LA class- B	5.0%	5.8%	10.8%	5.0%	0.23	0.90 to 0.94
Esophagitis LA class- C	0.8%	0.8%	1.7%	0.8%	1.00	0.92 to 0.95
Z-line irregularities	28.3%	26.7%	23.3%	28.3%	0.81	0.90 to 0.94
Hiatal hernia	40.0%	40.8%	39.2%	39.2%	0.99	0.96 to 0.97
vegetative / proliferative lesion	1.7%	1.7%	1.7%	1.7%	1.00	0.78 to 0.88
Esophageal varices	1.7%	2.5%	3.3%	1.7%	0.91	0.91 to 0.95

CI- confidence interval; * chi-square test

Table III. Results of the evaluation of the gastric antrum

Gastric antrum	Endoscopists				P value*	CI % for K value
	A	B	C	D		
normal	11.7%	11.7%	8.3%	11.7%	0.81	0.91 to 0.95
niche	4.2%	5.8%	5.8%	4.2%	0.90	0.95 to 0.97
atrophy	20.8%	18.3%	10.8%	20.8%	0.14	0.88 to 0.93
erythema	75.0%	79.2%	87.5%	74.2%	0.04	0.90 to 0.94
Polyp(s)	10.0%	6.7%	7.5%	10.0%	0.75	0.91 to 0.95
Intestinal metaplasia	7.5%	9.2%	14.2%	6.7%	0.19	0.88 to 0.93

CI- confidence interval; * chi-square test

In the assessment of gastric corpus, we obtained a very good agreement 0.9689 for evaluating normal gastric mucosa, intestinal metaplasia (0.9842) and angiodysplasia (0.9820). Also, interobserver variability was very good in assessment all of the lesions identified in the gastric corpus (Table IV).

In the assessment of duodenal bulb, we obtained very good agreement for evaluating ulcer niche (0.9842). The other questions found a very good agreement, and one question about atrophic mucosa in the second part of the duodenum being the most difficult to reach an agreement (0.8075) (Table V).

After assessing video-recordings from the second part of the duodenum we obtained very good agreement for evaluating normal mucosa (85.75%), erosions (80.75%) and again the question on atrophic mucosa in the second part of the duodenum being the most difficult to agree upon (58.64%). The interobserver variability in our study was moderate to very good in assessment of the lesions identified in the second part of the duodenum. (Table VI).

Discussion

This study had some limitations. First of all, this was a single-center study and the sample size may not be large enough. Secondly, given that this study analyzed previously obtained endoscopic video-recordings, it was difficult to evaluate as many details as possible in real time. Thirdly, magnifying endoscopy was additionally used, but not for every case, in our study and the results may have been influenced. In the situations where the endoscopic appearance was considered normal by the endoscopist, virtual chromoendoscopy or magnification was not used. If lesions were identified, for example polyps, gastric ulcers, areas of intestinal metaplasia, the endoscopist used for the morphological evaluation of the detected changes inspection in linked color imaging (LCI) and blue light imag-

Table IV. Results of the evaluation of the gastric corpus

Gastric corpus	Endoscopists				P value*	CI % for K value
	A	B	C	D		
normal	46.7%	46.7%	45.0%	45.8%	0.99	0.95 to 0.97
benign niche	1.7%	2.5%	3.3%	1.7%	0.91	091 to 0.95
atrophy	25%	25%	25%	25%	0.98	0.94 to 0.97
erythema	31.7%	35.8%	42.5%	31.7%	0.25	0.94 to 0.96
polyp(s)	9.2%	7.5%	8.3%	9.2%	0.98	0.90 to 0.94
Intestinal metaplasia	3.3%	3.3%	4.2%	3.3%	1.00	0.97 to 0.98
angiodysplasia	3.3%	3.3%	2.5%	3.3%	1.00	0.97 to 0.98

CI- confidence interval; * chi-square test

Table V. Results of the evaluation of the duodenal bulb

Duodenal bulb	Endoscopists				P value*	CI % for K value
	A	B	C	D		
normal	74.2%	74.2%	76.7%	73.3%	0.96	0.92 to 0.95
erosion	10.8%	10.0%	9.2%	10.8%	0.98	0.92 to 0.95
niche	3.3%	3.3%	4.2%	3.3%	1.00	0.97 to 0.98
atrophy	2.5%	2.5%	0.8%	1.7%	0.88	0.74 to 0.85
polyp(s)	5.0%	5.0%	3.3%	5.0%	0.94	0.96 to 0.98
ulcer scar	0.8%	1.7%	1.7%	0.8%	1.00	0.85 to 0.91

CI- confidence interval; * chi-square test

Table VI. Results of the evaluation of the second duodenum

D II	Endoscopists				P value*	CI % for K value
	A	B	C	D		
normal	92.5%	93.3%	95.8%	91.7%	0.66	0.81 to 0.89
erosion	2.5%	1.7%	0.8%	2.5%	0.88	0.74 to 0.85
atrophy	0.8%	1.7%	1.7%	0.0%	0.76	0.45 to 0.69

CI- confidence interval; * chi-square test ing (BLI) modes with a maximum optical magnification of 145 x which provided a highly detailed image of the mucosal surface and vascular patterns.

Prior studies on magnification endoscopy and minimal change esophagitis in non-erosive reflux disease patients showed substantial inter-observer agreement [6-9]. Finally, the quality of the image, any blurring image caused by the endoscopist's hand movement, lens fogging or poor cooperation may impair the results.

We found that variability is extensive in the assessment of images from upper endoscopy.

Similar results have been reported from other diagnostic disciplines, for example assessment of carotid plaques [10]. Variability among observers in our study ranged from moderate to very good with the highest level of agreement in answering questions regarding characteristic findings such as normal mucosa, esophagitis Class A Los Angeles, hiatal hernia for the esophagus endoscopic evaluation, benign ulcer niche in gastric antrum, normal gastric corpus mucosa, intestinal metaplasia and angiodysplasia in gastric corpus. The question on atrophic mucosa in the first and second part of the duodenum was the most difficult to agree upon.

In our study we have asked four endoscopists with different level of experience to complete the questionnaire after assessment of 120 video-recordings. Some studies [11-13] reveal that experience leads to a higher degree of agreement, while other do not [14, 15]. There have been studies in which live endoscopic video-recordings were presented that could have improved the degree of agreement, but there is a study in which live endoscopic images were used and the degree of agreement was not significantly modified [16, 17].

Interesting to note is the fact that Lundell et al. through their study found that greater experience did not lead to a higher degree of agreement [15].

Still images from gastroscopy fail to document motility which is just as significant as mucosal changes. Video-recording the entire examination may address these deficiencies, but for practical reasons, it is uncertain if video-recordings are a realistic way to systematically document gastroscopy. A standardized set of still images will always be the second best and more practical method. The ESGE has suggested a series of eight reference images for the documentation of upper endoscopic procedures [3]. In Figure 1 and Figure 2 are represented images of the gastric antrum and corpus according to the previously mentioned ESGE guideline.

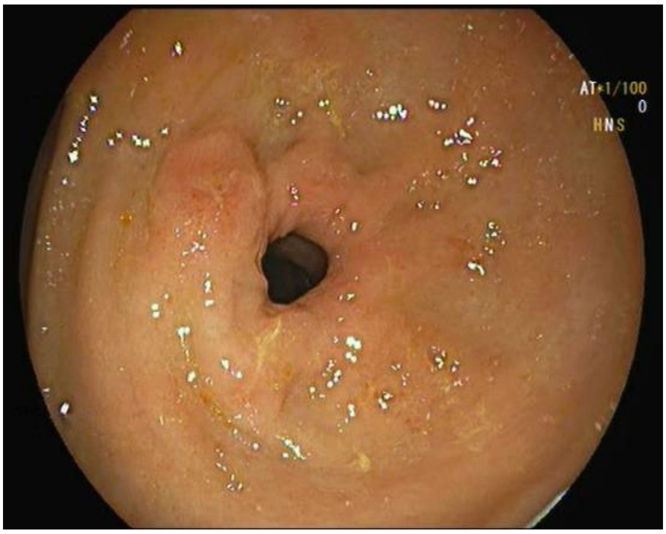

Fig. 1. Endoscopic aspect - antrum

Fig. 2. Endoscopic aspect - corpus

Conclusion

In summary, the present study found that the variability between observers in the assessment of images obtained from patients that underwent conventional upper gastrointestinal endoscopy in our center was acceptably good.

Acknowledgements

We are grateful to our colleagues who contributed to this study through their survey responses.

This work was supported by a grant of Ministry of Research and Innovation-project number ID P_34_498, within MFE 2014-2020-POC.

Authors' contribution

BI- substantial contributions to conception and design, performing the systematic literature research, selecting studies to be included, analysis and interpretation of data, drafting the article and revising it, final approval of the draft for publication.

VS – application of statistical techniques to analyze study data, creation and presentation of the published work, specifically writing the initial draft, review of the article, final approval of the draft for publication.

PM – collected the data, draft manuscript preparation, interpretation of data for the article, revising it critically for important intellectual content, final approval of the version to be published.

GG - collected the data, draft manuscript preparation, interpretation of data for the article, revising it critically for important intellectual content, final approval of the version to be published.

BM - interpretation of data, revising it critically for important intellectual content, final approval of the version to be published
CN – collected the data, draft manuscript preparation, interpretation of data for the article, revising it critically for important intellectual content, final approval of the version to be published

References

1. Idan Levy, MD, GI Fellow Professor, Ian M. Gralnek, MD, MSHS, FASGE, Head of Department; Complications of diagnostic colonoscopy, upper endoscopy, and enteroscopy; Best Pract Res Clin Gastroenterol. 2016 oct; 30(5):705-718.
2. Bendtsen F, Skovgaard LT, Sorensen TIet al. Agreement among multiple observers on endoscopic diagnosis of esophageal varices before bleeding. Hepatology 1990;11: 341-347.
3. Minimal standard terminology in digestive endoscopy. European Society of Gastrointestinal Endoscopy. Endoscopy 2000;32(2):162- 188.
4. Rey JF, Lambert R. ESGE recommendations for quality control in gastrointestinal endoscopy: guidelines for image documentation in upper and lower GI endoscopy. Endoscopy 2001;33(10):901-903.
5. Fleiss JL, Levin B, Paik MC (2003) Statistical methods for rates and proportions, 3rd ed. Hoboken: John Wiley & Sons.
6. Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991. pp 4039.
7. Hatlebakk JG. Endoscopy in gastro-esophageal reflux disease. Best practice & research Clinical gastroenterology. 2010;24(6):775-786.
8. Wasielica-Berger J, Kemona A, Kiśluk J, et al. The added value of magnifying endoscopy in diagnosing patients with certain gastresophageal reflux disease. Advances in medical sciences. 2018;63(2):359-366.
9. Robles-Medranda C, Valero M, Soria-Alcívar M, et al. Detection of minimal mucosal esophageal lesions in non-erosive gastresophageal reflux disease using optical enhancement plus optical magnification. Endoscopy International Open. 2019;7(8):979-986.
10. Amano Y, Ishimura N, Furuta K, et al. Interobserver agreement on classifying endoscopic diagnoses of nonerosive esophagitis. Endoscopy. 2006 Oct;38(10):1032-1035.
11. Lovett JK, Gallagher PJ, Rothwell PM. Reproducibility of histological assessment of carotid plaque: implications for studies of carotid imaging. Cerebrovasc Dis. 2004;18(2):117-123.
12. Orlandi F, Brunelli E, Feliciangeli G, et al. Observer agreement in endoscopic assessment of ulcerative colitis. Ital J Gastroenterol Hepatol. 1998;30(5)539-541.
13. Armstrong D, Bennett JR, Blum AL, et al. The endoscopic assessment of esophagitis: a progress report on observer agreement. Gastroenterology. 1996;111(1):85-92.
14. Pandolfino JE, Vakil NB, Kahrilas PJ. Comparison of inter- and intraobserver consistency for grading of esophagitis by expert and trainee endoscopists. Gastrointest Endosc. 2002; 56(5):639-643
15. Lundell LR, Dent J, Bennett JR, et al. Endoscopic assessment of esophagitis: clinical and functional correlates and further validation of the Los Angeles classification. Gut. 1999;45:172-180.
16. Bytzer P, Havelund T, Hansen JM. Interobserver variation in the endoscopic diagnosis of reflux esophagitis. Scand J Gastroenterol. 1993; 28(2):119-125
17. Anne M, Bjorn S, Eyvind J P. Impact of observer variability on the usefulness of endoscopic images for the documentation of upper gastrointestinal endoscopy. Scand J Gastroenterol. 2007 Sep;42(9):1106-1112.