Digestive and Liver Disease 46 (2014) 140–145

Contents lists available at ScienceDirect

Digestive and Liver Disease

journal homepage: www.elsevier.com/locate/dld

Digestive Endoscopy

Analysis of interobserver variability for endomicroscopy of the gastrointestinal tract

Arthur Hoffman[a,b,*], Johannes Wilhelm Rey[a,b], Lena Mueller[b], Torsten Hansen[c], Martin Goetz[b,d], Achim Tresch[e], Peter Robert Galle[b], Ralf Kiesslich[a,b]

[a] St. Mary's Hospital, Department of Medicine, Frankfurt, Germany
[b] Medical Department, Johannes Gutenberg University of Mainz, Germany
[c] Institute of Pathology, Johannes Gutenberg University Mainz, Germany
[d] Medical Department, University of Tübingen, Germany
[e] Max Planck Institute, Cologne, Germany

ARTICLE INFO

Article history:
Received 6 February 2013
Accepted 7 September 2013
Available online 7 November 2013

Keywords:
Colon cancer
Endomicroscopy
Experts
Inter observer variability
Pathologists
Students

ABSTRACT

Background: Endomicroscopy allows in vivo microscopic investigation of enteral mucosa during endoscopic examinations. The aim of this study was to determine interobserver variability in the evaluation of endomicroscopic pictures of several organs by groups of investigators composed of confocal experts, pathologists and students.
Methods: Twenty-five selected representative endomicroscopic pictures of the colon, stomach and oesophagus (total number, 75) were evaluated based on tissue, inflammatory and neoplastic changes (secondary endpoints). The endomicroscopic presence of neoplastic features was the primary endpoint and correlated with the final histological diagnosis.
Results: The kappa values for experts examining colon, stomach, and oesophagus pictures were 0.80, 0.91, and 0.488, respectively; for students 0.74, 0.684, and 0.527 and for pathologists 0.749, 0.633, and 0.346, respectively. Neoplasia was accurately diagnosed in 67–97% of patients with no significant differences between the 3 groups. Tissue differentiation was determined best by pathologists, whereas the degree of inflammation was better diagnosed by experts and students. In all 3 groups the diagnosis of oesophageal diseases was the most difficult.
Conclusions: Endomicroscopic images can be interpreted with high concordance. In our study, the diagnostic reliability was not different between students, endomicroscopic experts, and pathologists. Thus, endomicroscopy could be an additional and reliable imaging modality for diagnosing mucosal neoplasia of the gut.

1. Introduction

Modern endoscopy is triggered by advancing optical possibilities. The common goal of almost all new technologies is timely and accurate diagnosis of pre-malignant and early malignant lesions. This is of decisive importance for the prognosis of the patient. Ideally, one should be able to distinguish the type of tissue during endoscopy (neoplastic versus non-neoplastic). Preliminary steps in tissue differentiation paved the way for magnifying and high-resolution endoscopy. By observing the mucosal surface from a close distance and by its subsequent analysis, experts have been able not only to evaluate the architecture of colonic crypts [1,2], but also to predict their final histology.

Confocal endomicroscopy is a relatively new imaging modality of gastrointestinal endoscopy that allows in vivo microscopic investigation during an ongoing endoscopic examination. It allows in vivo histological investigation of the mucosal surface and subsurface at subcellular resolution. Cells, tissue and vascular structures can all be viewed along with the endoscopic image [3–6].

Endomicroscopy has proven clinical benefit and diagnostic accuracy for a wide range of diseases. However, the available endomicroscopic systems are expensive and the technique itself is considered to be highly examiner-dependent.

The aim of the present study was to analyse the interobserver variability and diagnostic accuracy rates among different groups of observers (all without prior endomicroscopic training), as to evaluate the baseline diagnostic characteristics of endomicroscopic images from 3 different organs.

* Corresponding author at: St. Marienkrankenhaus Frankfurt, Richard-Wagner-Str. 14, 60318 Frankfurt, Germany.
E-mail address: a.hoffman@katharina-kasper.de (A. Hoffman).

1590-8658/$36.00
http://dx.doi.org/10.1016/j.dld.2013.09.004

2. Methods

The database of endomicroscopic pictures of the interdisciplinary endoscopy at the First Department of the University Clinic of Medicine in Mainz was used to select appropriate endomicroscopic images for the study.

The database consisted of 134,503 pictures (up to the year 2010), from which 25 representative endomicroscopic images of the stomach, oesophagus, and colon (75 images total) were "hand-selected". Only endomicroscopic pictures of the Pentax-System (Pentax EC3870K, Japan) were used.

The selected pictures had been taken between the years 2003 and 2008, and their corresponding histology was known and confirmed by 2 independent pathologists. All the selected pictures were assessed with regard of image quality, sharpness, and light exposure and were graded excellent by a single and very experienced investigator (RK).

An online questionnaire was developed, with questions concerning inflammation, type of tissue, grading, and final diagnosis (neoplasia yes/no).

Endomicroscopic experts were selected based on their experience of more than 200 endomicroscopic examinations. The control group consisted of medical students, with no experience in endomicroscopy, who had just passed their histological training at the Johannes Gutenberg University of Mainz, Germany.

The pathologists involved in this study were experts in the field of gastrointestinal pathology, but not in the field of confocal imaging.

The observers involved did not undergo structured training before starting the online questionnaire evaluating their response to endomicroscopic images.

The primary endpoint of the study was to estimate the interobserver variability in evaluating the presence of neoplasia. Secondary endpoints were grading of inflammation and dysplasia.

The questionnaire was published online via a website designed with Adobe Dreamweaver. All graphic diagrams were processed with Adobe Photoshop. Programming languages were HTML and Java.

The complete questionnaire was sent to all study participants via the corresponding link to the website (see supplementary Figure S1). The time for answering the questions was not restricted. However, the questionnaire could not be changed after the answers were given.

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.dld.2013.09.004.

2.1. Sample size calculation

To estimate concurrence between groups and thus establish the reproducibility of the diagnoses, we used the kappa value [11,13,14].

Sample size was calculated using a two-sided 95% confidence interval for an anticipated kappa value of 0.8, with a breadth of 0.15, at a disease rate of 1/3; this yielded a case count of 70. Sensitivity, specificity, and confidence intervals were calculated, and reliability was investigated and evaluated using kappa statistics.

For statistical evaluation we used both the statistics programme SPSS (version16.0) and Excel (version 2008).

The answers to the questionnaires were tabulated and then evaluated using the predictive analytics software programme of SPSS Inc.

3. Results

The primary aim of the present study was to determine the interobserver variability in the evaluation of endomicroscopic pictures. The identification of malignant lesions (neoplasia present yes/no) was the main outcome of the study (see Tables 1–3 and Figs. 1–3).

Furthermore, tissue characterisation, and grading of inflammation and neoplasia were secondary outcome measurements (as specified in the questionnaire).

Tables 1–3 show the kappa values for the 3 groups (students, experts and pathologists) and the sensitivity, specificity, positive and negative predicting values for the endpoint "presence of neoplasia yes/no".

Furthermore, Figs. 1–3 show the number of correct diagnoses for each group and each parameter.

3.1. Primary and secondary endpoints

3.1.1. Colon

The presence of neoplasia could best be diagnosed within the colon, where experts, pathologists, and students showed similar rates of correct answers (88%, 91%, and 88%, respectively).

However, the type of the tissue could best be evaluated by pathologists, the inflammation was best graded by students and grading of neoplasia was best achieved by pathologists (see Fig. 1). The experts obtained the highest kappa values (0.8), followed by pathologists (0.749), and students (0.74). However, all values were within a very close and similar range.

3.1.2. Stomach

The primary endpoint (presence of neoplasia) was best achieved by pathologists (96%), followed by students (84%), and experts (83%).

Evaluation of the secondary endpoints was also markedly different. Tissue type was best diagnosed by pathologists (correct in 88%), compared to 60% for experts and 45% for students.

Corresponding Kappa values were: 0.91 for experts, 0.68 for students, and 0.633 for pathologists.

Table 1
Results of the experts for neoplasia in colon, stomach and oesophagus.

	Kappa	Kappa interval	Specificity	Sensitivity	Positive predictive value	Negative predictive value
Colon	0.8	[0.66–0.94]	97.8%	80.0%	96.0%	88.0%
Stomach	0.91	[0.81–1.00]	95.8%	96.3%	92.9%	97.9%
Oesophagus	0.488	[0.3–0.68]	70.8%	81.5%	61.1%	87.2%

Table 2
Results of the students for neoplasia in colon, stomach and oesophagus.

	Kappa	Kappa interval	Specificity	Sensitivity	Positive predictive value	Negative predictive value
Colon	0.74	[0.59–0.9]	97.8%	73.3%	95.7%	84.6%
Stomach	0.684	[0.53–0.84]	75.0%	100%	69.2%	100%
Oesophagus	0.527	[0.33–0.72]	77.1%	77.8%	65.6%	86.0%