Digestive and Liver Disease 46 (2014) 518–522

Contents lists available at ScienceDirect

Digestive and Liver Disease

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

Digestive Endoscopy

# Interobserver agreement for evaluation of imaging with single operator choledochoscopy: What are we looking at?

Amrita Sethi [a], Jessica Widmer [b], Neeral L. Shah [c], Douglas K. Pleskow [d], Steven A. Edmundowicz [e], Divyesh V. Sejpal [f], Frank G. Gress [a], George H. Pop [b], Monica Gaidhane [b], Bryan G. Sauer [c], Peter D. Stevens [a], Michel Kahaleh [b,*]

[a] *Division of Digestive and Liver Disease, Columbia University Medical Center-NYPH, New York, NY, United States*
[b] *Division of Gastroenterology and Hepatology, Weill Cornell Medical College, New York, NY, United States*
[c] *Division of Gastroenterology and Hepatology, University of Virginia, Charlottesville, VA, United States*
[d] *Division of Gastroenterology, Beth Israel Deaconess Medical Center, Boston, MA, United States*
[e] *Division of Gastroenterology, Washington University School of Medicine, St. Louis, MO, United States*
[f] *Division of Gastroenterology, North Shore LIJ Health System, Manhasset, NY, United States*

ARTICLE INFO

*Article history:*
Received 28 May 2013
Accepted 7 February 2014
Available online 16 March 2014

*Keywords:*
Cholangioscopy
Choledochoscopy
SOC
SpyGlass

ABSTRACT

*Background:* Single operator choledochoscopy is a platform used to assist in the confirmation of diagnosis of biliary lesions. However, there are little data regarding the interobserver agreement of imaging interpretation. Our objective was to assess the interobserver agreement in single operator choledochoscopy interpretation.

*Methods:* 38 De-identified SPY Choledochoscopy video clips were sent to 7 interventional endoscopists. They were asked to score the videos on presence of four criteria selected by the investigators: growth, stricture, hyperplasia, and ulceration. Observers also chose a final diagnosis from the categories of cancer, hyperplasia, inflammation, or normal. Kappa scores were calculated for the scoring of the four criteria and for the selection of the final diagnosis.

*Results:* The overall interobserver agreement was fair in scoring for the presence of a growth ($K = 0.28$, SE 0.035) and stricture ($K = 0.32$, SE 0.035). Scoring for ulceration was slight to fair ($K = 0.17$, SE 0.035). There was only slight agreement for the presence of hyperplasia ($K = 0.11$, SE 0.035); and presumed final diagnosis based on imaging ($K = 0.18$, SE 0.022).

*Conclusion:* The results of this study support the need for an effort to identify and validate cholangioscopy imaging criteria for biliary pathology. This may assist in improving the reliability of the diagnostic value of cholangioscopy as its use becomes more widespread.



## 1. Introduction

Single operator cholangioscopy systems represent a fairly recent advance in the field of peroral cholangioscopy, which were first described in the 1970s. In the United States, the only such system currently available is SpyGlass direct visualization (Boston Scientific, Natick, MA). It provides four-way steerability, dedicated irrigation channels and a 1.2 mm working channel through which diagnostic and therapeutic devices can be used. These instrumentation advantages and its high image quality have led to renewed interest in the technique with subsequent expanded use [1,2]. Currently, the most common indications for cholangioscopy are stone therapy and evaluation of indeterminate biliary strictures [1,3]. Less common indications include guidewire placement during endoscopic retrograde cholangiopancreatography (ERCP), assessment of post-liver transplantation biliary strictures, and evaluation of indeterminate intraductal filling defects or irregularities of the bile duct seen on imaging studies such as computed tomography (CT), magnetic resonance imaging (MRI), endoscopic ultrasound (EUS) or ERCP [1,4,5]. Rare applications include staging and ablation of biliary neoplasms, investigation of recurrent pancreatitis, and evaluation of hemobilia [1,6].

In 2011, results from the largest published registry showed that the sensitivity and specificity of SpyGlass visual diagnosis

* Corresponding author at: Division of Gastroenterology & Hepatology, Weill Cornell Medical College, New York, NY 10021, United States. Tel.: +1 646 962 4000; fax: +1 646 962 0110.

*E-mail address:* mkahaleh@gmail.com (M. Kahaleh).

http://dx.doi.org/10.1016/j.dld.2014.02.004
1590-8658/

for malignancy were 78% and 82%, respectively [3]. While reporting on diagnosis by visual impression, however, no definitions of criteria for malignancy were provided nor seem to be uniformly employed by the multiple centres during this study. Later publications also reported diagnostic accuracies of visual impression by cholangioscopy, however, were based on definitions provided by the investigators rather than reference to standardized, validated definitions [7,8]. Additional cholangioscopic platforms, such as ultra-slim endoscopes and electronic cholangioscopes appear to provide improved quality images. However, no formal comparison has been made between the systems. In fact, no studies have been reported regarding the interobserver agreement of cholangioscopic imaging and furthermore, no validation studies have been reported regarding definitions of cholangiosopic imaging criteria for the diagnosis of biliary and pancreatic lesions. The objective of this pilot multicenter study was to assess the interobserver agreement and variance in interpretation of imaging alone by SOC.

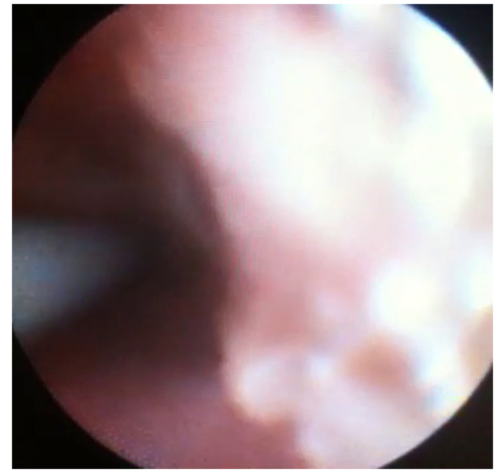

**Fig. 1.** Choledochoscopy image showing growth in the duct.

## 2. Methods

Thirty-eight de-identified SpyGlass choledochoscopy video clips taken during exams performed at the University of Virginia between 2008 and 2010 were compiled and sent out to seven selected interventional endoscopists (AS, NS, DP, SE, DV, FG, PDS). The reviewers were considered experts in the field based on having performed >50 cholangioscopies. These procedures were performed by one expert choledochoscopist (MK) in a single centre over a two year time period. All videos had been recorded in a .MKT and .MP4 format. Videos were selected from a library of videos for which there was an available corresponding final diagnosis. Videos were included if they were devoid of images that would bias the diagnosis, such as manoeuvres with forceps, or stone fragments. The videos were less than two minutes in duration and were compiled of images of intraductal strictures and lesions. The clips were not selected based on video quality as determined by the endoscopist. There were no annotations or labelling included in the video clips. In order to ellicit an interpretation based soley on visual findings, and not biased by background information, the reviewers were blinded to all clinical information related to the indication of the choledoscopy and the final diagnosis. No data regarding patient characteristics were collected. No fluoroscopic or endoscopic images were provided to the reviewers.

**Fig. 2.** Choledochoscopy image showing stricture in the duct.

Given the lack of published criteria at the time this study was devised, four visual features were identified that had been uniformly used by the performing endoscopist during the time of his exams, which allowed for a self-assessed accuracy of 90% in diagnosing malignancy by visualization. These impressions of malignant vs benign lesions and individual features were recorded by the performing endoscopist at the time of the actual exam during which the video clip was recorded, and referenced at the time of the study when determining the final diagnosis. Reviewers were asked to score the videos on the presence and severity of these four features: growth, stricture, hyperplasia, and ulceration (Figs. 1–4). Given lack of set definitions in the literature at the time this study was performed, standard reference images were not sent to the reviewers. Observers were also asked to select a final diagnosis from the categories of cancer, hyperplasia, inflammation, or normal. Lastly they were asked to grade the quality of the videos as poor, fair, and good, again based on their interpretation; no reference library was provided. Accuracy, in diagnosing benign versus malignant lesions was measured by comparing the observer's final diagnosis to the final tissue diagnosis (by biopsy, brushing, or surgical specimen) or clinical follow-up for >6 months. Patients were considered to have benign disease if they were alive after 6 months from their exam with no diagnosis of malignancy. A diagnosis of cancer was considered malignant, whereas hyperplasia, inflammation, and normal where considered benign. University of Virginia HSR-IRB approval was obtained on 11/12/2010 (Protocol no. 15335).

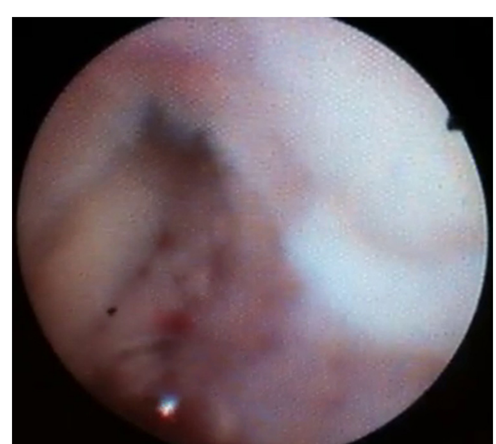

**Fig. 3.** Choledochoscopy image showing hyperplasia in the duct.