Interobserver agreement of EUS elastography in the evaluation of solid pancreatic lesions

João-Bruno Soares¹, Julio Iglesias-Garcia², Bruno Gonçalves¹, Björn Lindkvist³⁴, José Lariño-Noia³ª, Pedro Bastos¹, Ana Célia Caetano¹⁶, Aníbal Ferreira¹⁶, Pedro Pimentel-Nunes⁷⁸, Luís Lopes⁶⁹, Pedro Moutinho-Ribeiro¹⁰, J. Enrique Dominguez-Muñoz²ª

¹Department of Gastroenterology, Hospital of Braga, Braga, Portugal, ²Department of Gastroenterology, University Hospital of Santiago de Compostela, Santiago de Compostela, A Coruña, ³Foundation for Research in Digestive Diseases, University Hospital of Santiago de Compostela, Santiago de Compostela, Spain, ⁴Department of Gastroenterology, Sahlgrenska University Hospital, ⁵Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, ⁶Surgical Sciences Research Domain, Life and Health Sciences Research Institute, School of Health Sciences, University of Minho, Braga, ⁷Department of Gastroenterology, Portuguese Oncology Institute, Oporto, ⁸Department of Gastroenterology, Hospital Center of Alto Minho, Viana do Castelo, ⁹Department of Gastroenterology, Hospital Center of Alto Ave, Guimarães, Portugal

ABSTRACT

Background and Objectives: Previous reports assessing the reproducibility of endoscopic ultrasound elastography (EUS-E) in evaluation of solid pancreatic lesions (SPL) involved only experienced endosonographers. We aimed to assess the interobserver agreement (IOA) of EUS-E in the evaluation of SPL by endoscopists with different levels of experience in EUS and EUS-E. Materials and Methods: A cross-sectional observational multicenter study was designed and included 11 endoscopists who were divided into four groups: Group A (long experience in EUS and EUS-E); Group B (short experience in EUS and EUS-E); Group C (long experience in EUS and no experience in EUS-E); and Group D (no experience in EUS or EUS-E). The observers independently classified the patterns of 60 video sequences of EUS-E, after a 20-min training session. For each group, we calculated IOA (kappa statistic, κ) of EUS-E and the diagnostic accuracy of EUS-E for pancreatic malignancy, by comparing the pattern of EUS-E indicative of malignancy (heterogeneous or homogenous blue) with the final diagnosis. Results: The overall IOA was moderate (κ = 0.42; 95% confidence interval (CI) 0.33-0.52). The IOA of Group A (κ = 0.80; 95% CI 0.65-1.00) was significantly higher than that of Groups B (κ = 0.54; 95%CI 0.40-0.71), C (κ = 0.54; 95%CI 0.39-0.68), and D (κ = 0.28; 95%CI 0.14-0.40). IOA of Groups B and C was not significantly different, but it was significantly higher than that of Group D. The diagnostic accuracy of Group A (area under the curve under summary receiver operating characteristic (AUROC) = 0.83; 95%CI 0.75-0.90) was not significantly different from that of Group B (AUROC = 0.77; 95%CI 0.71-0.83), but it was significantly higher than that of Group C (AUROC = 0.74; 95%CI 0.67-0.81) and D (AUROC = 0.74; 95%CI 0.67-0.81). No significant difference was seen between Groups B, C, and D for diagnostic accuracy. Conclusion: EUS-E is reproducible in the evaluation of SPL, even between endoscopists with no or limited experience in EUS and/or EUS-E. Reproducibility and diagnostic accuracy increase with experience in EUS and EUS-E.

Key words: Diagnostic accuracy, endoscopic ultrasound elastography, interobserver agreement, reproducibility, solid pancreatic lesions

INTRODUCTION

Endoscopic ultrasonography (EUS) provides high-resolution images of the pancreas and it is considered one of the most accurate methods for the diagnosis and staging of chronic inflammatory, cystic,
Elastography is a method for the real-time evaluation of tissue stiffness, which has been used for the analysis of superficial organ lesions, such as those of the breast and prostate. Pressure is applied to the lesion during the exploration, and resulting differences in distortion between hard and soft tissues are used for the real-time analysis of their stiffness. Several recent studies have shown EUS-E as a promising technique with a high accuracy for the differential diagnosis of solid pancreatic lesions (SPLs). Actually, features of the elastographic pattern in terms of homogeneity or heterogeneity, and predominant color, closely correlate with the histological features of the lesion. In a recent meta-analysis; the pooled sensitivity, specificity, and diagnostic odds ratio of EUS-E distinguishing benign from malignant solid pancreatic masses were 0.95 (95% confidence interval (CI) 0.94-0.97), 0.67 (95% CI 0.61-0.73), and 42.28 (95% CI 26.90-66.46), respectively.

Although EUS-E seems to be promising for evaluation of SPL, it is not clear whether the interpretation of EUS-E is reproducible among different endosonographers. Previous reports assessing the reproducibility of EUS-E for evaluation of SPL involved only experienced endosonographers. The main aim of this study was to assess the interobserver agreement (IOA) of EUS-E in the evaluation of SPL by endoscopists with different levels of experience in EUS and EUS-E. We additionally evaluated the accuracy of EUS-E for the diagnosis of pancreatic malignancy by endoscopists with different levels of experience in EUS and EUS-E.

### MATERIALS AND METHODS

#### Design of the study and selection of patients

This was a cross-sectional observational study with two aims. The primary aim was to assess the IOA of EUS-E in the evaluation of SPL by endoscopists with different levels of experience in EUS and EUS-E. The secondary aim was to assess the diagnostic accuracy of EUS-E for the diagnosis of pancreatic malignancy by endoscopists with different levels of experience in EUS and EUS-E.

A total of 60 patients with SPLs who underwent routine EUS at the Department of Gastroenterology, University Hospital of Santiago de Compostela (Spain) during 2011 were consecutively included in this study after giving informed consent for EUS. A final diagnosis of malignant or benign tumor was defined according to the following reference methods:

1. Histological findings of surgical specimens in patients undergoing surgery;
2. Cytological findings definitely positive for malignancy together with compatible EUS and computed tomography (CT) findings for final diagnosis of malignant disease, in patients with unresectable tumors, and
3. EUS and CT findings at entry, and a minimum follow-up period of 6 months including EUS-FNA and CT, for a final diagnosis of benign disease in patients with benign cytological findings.

All of the material provided for the study was anonymous, and in no instance was a patients’ identity revealed. The study was approved by the local institutional review board and conducted in accordance with the Declaration of Helsinki and its amendments, and Good Clinical Practice guidelines.

#### Technique of EUS and selection of videos

In each of the 60 patients, the SPL was evaluated with standard EUS imaging and EUS-E. EUS was performed with a linear EUS probe (EG3830UTK; Pentax Europe GmbH, Hamburg, Germany) attached to a platform Preirus (Hitachi Medical Systems GmbH, Wiesbaden, Germany), which includes the elastography module. All procedures were done by two experienced endosonographers. The complete description of the technique of EUS-E has been reported elsewhere. For each patient, one video sequence was recorded for 30 s [Figure 1]. Each video sequence also included a B-mode standard EUS image of the lesion of interest.
Each video sequence was labeled with a random number by an endosonographer who had not participated in the EUS procedure and was blinded to the clinical history and the pathological diagnosis. The observers were provided with a pen drive containing the 60 video sequences and were allowed unlimited time to review the videos. On the other hand, the observers were blinded to the clinical history and the pathologic diagnosis and to each others evaluation. No prior selection was made based on the quality of recorded images to avoid inducing any bias in the IOA evaluation.

Selection of observers and evaluation of videos
A total of 11 endoscopists from six European EUS centers participated in this study. They were divided into four groups according to their experience in EUS and EUS-E:

- Group A included two endosonographers with long experience in EUS (>1,000 procedures) and EUS-E (>200 procedures);
- Group B included three endosonographers with a 3-months’ experience in EUS (>100 procedures) and EUS-E (>20 procedures);
- Group C included three endosonographers with long experience in EUS (>1,000 procedures), but no experience in EUS-E;
- Group D included three endoscopists with no experience in EUS or EUS-E.

A kick off session of 20 min was undertaken to share the principles of the techniques and to make everybody acquainted with the parameters of EUS-E under evaluation.

Observers were asked to classify the elastographic pattern of the lesion of interest into four types according to the predominant color and the homogeneity or heterogeneity of color distribution (homogeneous green = 1, heterogeneous green = 2, heterogeneous blue = 3, homogeneous blue = 4).[17]

We also calculated the diagnostic accuracy for pancreatic malignancy, by comparing the patterns of EUS-E indicative of pancreatic malignancy with the final diagnosis as previously described. We considered heterogeneous or homogeneous blue patterns in EUS-E as indicative of pancreatic malignancy.

Statistical analysis
The Fleiss kappa (κ) statistic was used to evaluate the IOA among observers. An individual κ for each group of observers, as well as an overall κ, was determined for EUS-E. The κ-values were interpreted according to the guidelines proposed by Landis and Koch.[19] The κ statistics allocates a score of zero if the agreement is no better than would be expected by chance, whereas perfect agreement is indicated by a kappa value of 1. Scores can also be negative if there is consistent disagreement. In detail, κ-values from 0.00 to 0.19 represent slight agreement, 0.20 to 0.39 fair agreement, 0.40 to 0.59 moderate agreement, 0.60 to 0.79 substantial agreement, and more than 0.80 is considered almost perfect agreement. The κ-values were considered statistically significant when 95% CI of κ-values was superior to 0. Bootstrap resampling was used to calculate the 95 % CI of κ-values. Statistical comparison of κ-values between groups was done using κ analysis extension for ArcView 3.2. We also evaluated the sensitivity, specificity, positive predictive value, negative predictive value, and area under the curve under summary receiver operating characteristic (AUROC) of each group and modality for the final diagnosis of pancreatic malignancy by using the heterogeneous or homogeneous blue pattern in EUS-E as indicative of the presence of pancreatic malignancy.[18,20]

With the exception of comparison of κ-values (see above), all statistical analyses were performed using the software Statistical Package for Social Sciences (SPSS) 18.0 (IBM SPSS Statistics, Chicago, Illinois, USA).

RESULTS

Patients’ characteristics
A total of 60 patients (17 females and 43 males with a mean age of 64 ± 15 years) were included in the study.
Soares, et al.: IOA of EUS-E for SPL

[Table 1]. The mean size of pancreatic masses was $36.5 \pm 15.9$ mm. The lesions were located mostly in the pancreatic head. The diagnosis was based on EUS-FNA in 43 patients, on endoscopic ultrasonography-fine needle biopsy (EUS-FNB) in 14 patients, on surgery in one patient, and on follow-up in two patients. As determined according to the reference methods, the final diagnoses were as follows: Pancreatic adenocarcinoma (45 patients); inflammatory mass in the context of chronic pancreatitis (10 patients); pancreatic neuroendocrine tumor (three patients); autoimmune pancreatitis (one patient); and colon cancer metastasis (one patient).

**IOA**

The IOA evaluation data are presented in Table 2. The overall IOA for EUS-E was moderate. Group A had the highest IOA followed by Groups B, C, and D. IOA of Group A was significantly higher than that of Groups B, C, and D. IOA of Groups B and C was not significantly different, but it was significantly higher than that of group D.

**Diagnostic accuracy for pancreatic malignancy**

The diagnostic accuracy data for pancreatic malignancy are presented in Table 3. Group A had the highest accuracy for diagnosis of pancreatic malignancy followed by Groups B, C, and D. Diagnostic accuracy of Group A was not significantly different from that of Group B, but it was significantly higher than that of Groups C and D. No significant difference was seen between Groups B, C, and D, in terms of diagnostic accuracy.

**DISCUSSION**

This is the first study evaluating the IOA of EUS-E in the evaluation of SPLs by endoscopists with different levels of experience in EUS and EUS-E. Our data suggest that EUS-E is reproducible in the evaluation of SPL even between endoscopists with no/low experience in EUS and/or EUS-E. Our data also suggest that experience in both EUS and EUS-E influences the IOA. This is based on the observation that the IOA for EUS-E of group A (observers with long experience in EUS and EUS-E) was significantly higher than that of all the other groups. The fact that Groups B (observers with short experience in EUS and EUS-E) and C (observers with long experience in EUS but not significantly different, but it was significantly higher than that of group D.}

Table 1. Characteristics of patients and pancreatic lesions included in an analysis of interobserver agreement of endoscopic ultrasound elastography in the evaluation of solid pancreatic lesions

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of patients, mean ± SD (years)</td>
<td>64 ± 15</td>
</tr>
<tr>
<td>Gender of patients (females:Males)</td>
<td>17:43</td>
</tr>
<tr>
<td>Lesion size, mean ± SD, mm</td>
<td>36.5 ± 15.9</td>
</tr>
<tr>
<td>Location of lesions in pancreas (head/body/tail)</td>
<td>43/15/2</td>
</tr>
<tr>
<td>Final diagnosis</td>
<td></td>
</tr>
<tr>
<td>Pancreatic adenocarcinoma</td>
<td>45</td>
</tr>
<tr>
<td>Inflammatory mass in the context of chronic pancreatitis</td>
<td>10</td>
</tr>
<tr>
<td>Pancreatic neuroendocrine tumor</td>
<td>3</td>
</tr>
<tr>
<td>Autoimmune pancreatitis</td>
<td>1</td>
</tr>
<tr>
<td>Metastatic colon cancer metastasis</td>
<td>1</td>
</tr>
</tbody>
</table>

SD: Standard deviation

Table 2. Results of interobserver agreement for each group of endosonographers

<table>
<thead>
<tr>
<th>Group</th>
<th>Interpretation of agreement</th>
<th>k (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (2 observers with long experience in EUS and EUS-E)</td>
<td>Almost perfect</td>
<td>0.80 (0.65-1.00)</td>
</tr>
<tr>
<td>B (3 observers with short experience in EUS and EUS-E)</td>
<td>Moderate</td>
<td>0.54 (0.40-0.71)*</td>
</tr>
<tr>
<td>C (3 observers with long experience in EUS and no experience in EUS-E)</td>
<td>Moderate</td>
<td>0.54 (0.39-0.68)*</td>
</tr>
<tr>
<td>D (3 observers with no experience in EUS or EUS-E)</td>
<td>Fair</td>
<td>0.28 (0.14-0.40)*</td>
</tr>
<tr>
<td>Overall (11 observers)</td>
<td>Moderate</td>
<td>0.42 (0.33-0.52)</td>
</tr>
</tbody>
</table>

κ: kappa value, 95%CI = 95% confidence interval, EUS: Endoscopic ultrasonography; EUS-E: Endoscopic ultrasonography elastography; Data are shown as κ (95% CI); †P < 0.05 vs. group A, ‡P < 0.05 vs. group D

Table 3. Diagnostic accuracy of each group of endosonographers for pancreatic malignancy

<table>
<thead>
<tr>
<th>Group</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
<th>AUROC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>92.9 (85.8-97.1)</td>
<td>78.9 (71.4-85.2)</td>
<td>70.1 (62.0-77.3)</td>
<td>78.9 (71.4-85.2)</td>
<td>0.83 (0.75-0.90)</td>
</tr>
<tr>
<td>Group B</td>
<td>72.7 (49.8-89.2)</td>
<td>75.8 (57.7-88.9)</td>
<td>78.8 (61.1-91.0)</td>
<td>69.7 (51.3-84.4)</td>
<td>0.77 (0.71-0.83)</td>
</tr>
<tr>
<td>Group C</td>
<td>93.8 (87.0-97.7)</td>
<td>93.5 (87.7-97.2)</td>
<td>93.6 (87.3-97.4)</td>
<td>92.1 (85.9-96.1)</td>
<td>0.74 (0.67-0.81)*</td>
</tr>
<tr>
<td>Group D</td>
<td>69.6 (47.1-86.7)</td>
<td>44.6 (31.3-58.5)</td>
<td>37.1 (25.9-49.5)</td>
<td>42.6 (29.2-56.8)</td>
<td>0.74 (0.67-0.81)*</td>
</tr>
</tbody>
</table>

95% CI = 95% confidence interval, PPV: Positive predictive value; NPV: Negative predictive value; AUROC: Area under the curve under summary receiver operating characteristic; Results of evaluation of 45 pancreatic adenocarcinoma and 15 nonpancreatic adenocarcinoma lesions (inflammatory mass in the context of chronic pancreatitis, n = 10; neuroendocrine tumor, n = 3; autoimmune pancreatitis, n = 1; metastasis, n = 1) were included for analysis. Pancreatic malignancy was defined by the presence of heterogeneous or homogeneous blue pattern in endoscopic ultrasound elastography (EUS-E). *P < 0.05 vs. Group A
Soares, et al.: IOA of EUS-E for SPL

no experience in EUS-E) had similar IOA for EUS-E suggests that long experience in EUS may influence the IOA of EUS-E and compensate for the lack of experience in EUS-E.

IOA for evaluation of SPL by EUS-E has been reported by some studies, but none has compared endosonographers with different levels of experience in EUS and EUS-E. Iglesias-Garcia et al., evaluated the IOA of EUS-E in 130 SPL by two EUS- and EUS-E-experienced endosonographers, obtaining a $\kappa$-value for elastographic patterns of 0.77 (95% CI 0.65-0.89). Giovannini et al., evaluated the IOA of EUS-E in 15 SPL by five EUS- and EUS-E-experienced endosonographers, reporting a $\kappa$-value for elastographic patterns of 0.524. Saffioti et al., evaluated the IOA of EUS-E in 258 SPL by five EUS- and EUS-E-experienced endosonographers, reaching a $\kappa$-value for elastographic patterns of 0.72. Taken together, these studies report a $\kappa$-value for EUS-E in the evaluation of SPL by experienced endosonographers that ranges from 0.52 to 0.77. In our study, the $\kappa$ was slightly higher (0.80) than previously reported.

This is also the first study comparing the diagnostic accuracy of EUS-E for pancreatic malignancy by endoscopists with different levels of experience in EUS and EUS-E. The results of the diagnostic accuracy evaluation were very similar to those of IOA evaluation. As for IOA, our data also suggest that experience in both EUS and EUS-E influences the diagnostic accuracy. This is based on the observation that diagnostic accuracy of EUS-E in Group A was significantly higher than in Groups C and D (observers with no experience in EUS or EUS-E). The similar diagnostic accuracy of EUS-E between Groups A and B suggests that short experience in EUS-E is sufficient to reach the diagnostic accuracy of experts. On the other hand, the similar diagnostic accuracy of EUS-E between Groups B and C suggests that long experience in EUS may influence the diagnostic accuracy of EUS-E and compensate for the lack of experience in EUS-E.

A limitation of EUS-E is that the qualitative image analysis performed is amenable to subjective interpretation of findings. Thus, methods for quantitative assessment of EUS-E have been recently developed, such as the hue histogram and the strain ratio. Although these methods have already proved to be helpful, there are still some limitations. All are based on computed automated analysis of regions of interest that are selected subjectively; thus, allowing the generation of selection bias. Moreover they have not yet been proved to be superior to qualitative analysis.

This study has some weaknesses. First, the small sample size, the low number of observers per group, and the low rate of nonneoplastic lesions (although similar to that of clinical practice) could influence our data. In fact, a major limitation of the study is the smaller number of observers in Group A (two observers) compared with other groups (three observers per group). This was the result of the low number of available experts on EUS-E and may have led to overestimation of IOA of Group A, compromising the comparison between this group and the other groups. Even so, this does compromise neither the overall data for IOA of EUS-E, nor the comparison between the other three groups, and thus does not influence the major conclusions of the study. Second, the procedures were done by the same experts who later evaluated the lesions (Group A). To reduce this bias, each video sequence was labeled with a random number and the experts blindly evaluated the lesions. Third, each video sequence also included a B-mode standard EUS image. Although this is similar to the clinical practice of EUS-E, evaluation of B-mode standard EUS image could influence evaluation of EUS-E images; thus, possibly contributing to overestimation of IOA and diagnostic accuracy of observers with long experience in EUS (Groups A and C). It has been shown that B-mode standard EUS may show changes strongly consistent with malignant tumor in 30% of solid pancreatic lesions (SPL). Finally, as we did not assess the intraobserver agreement of EUS-E, we cannot make any conclusions about intraobserver reproducibility of EUS-E.

In conclusion, we present the first multicenter study comparing the IOA and diagnostic accuracy of EUS-E in the evaluation of SPL by endoscopists with different levels of experience in EUS and EUS-E. Our data suggest that EUS-E is reproducible, even in groups with no or limited experience in these techniques. This study also hints that long experience in both EUS and EUS-E is a major contributor to the IOA and diagnostic accuracy of EUS-E. Nonetheless, these data should be submitted to external validation in larger studies.

REFERENCES


