Strain Assessment in Surgically Resected Inflammatory and Neoplastic Bowel Lesions

Bestimmung des Strain bei chirurgisch-resezierten entzündlichen und neoplastischen Darmläsionen

Abstract

Purpose: To investigate whether ultrasound-based strain imaging can discriminate between colorectal adenocarcinomas and stenotic Crohn's lesions in newly resected surgical specimens.

Materials and Methods: Resected surgical specimens from 27 patients electively operated for colorectal tumors or stenotic lesions from Crohn's disease were prospectively examined with ultrasonography using a Hitachi HV 900 US scanner with real-time elastography (RTE). Three different methods were applied to assess tissue strain: A four-level categorical visual classification, a continuous visual analog scale (VAS, 0 – 100) and a strain ratio (SR) measurement between the lesion and surrounding reference tissue. The imaged sections were marked and subsequently examined by a pathologist. Results from RTE were evaluated according to diagnosis, degree of fibrosis, inflammatory parameters, tumor stage and grade.

Results: 16 sections from Crohn's lesions, 18 sections from adenocarcinomas and 4 sections from adenomas were examined. Both adenocarcinomas and Crohn's lesions were found to be harder than the surrounding tissue, but they could not be discriminated from each other by any of the strain imaging evaluation methods. All adenocarcinomas had significantly higher strain ratios than adenomas. The categorical classification differentiated poorly between Crohn's lesions, adenocarcinomas and adenomas. Categorical evaluation and VAS score showed fair interobserver agreement. SR measurements provided semi-quantitative strain data and added improved information about elasticity properties, despite substantial intra-observer variation.

Conclusion: Sonoelastography with SR measurements and visual evaluation of strain differences could not differentiate stenotic Crohn's lesions from adenocarcinomas in resected bowel specimens.

Zusammenfassung


Introduction

Differentiation between benign and malignant tissue is a major challenge in medical imaging. Tissue samples are usually necessary for a reliable diagnosis, but they may be difficult to obtain. Pathological changes may alter tissue structure and hence its elastic properties. New imaging modalities have been introduced as potential tools for the discrimination between malignant lesions and benign tissue based on changes in elasticity or strain distribution patterns [1]. Real-time elastography (RTE) is an ultrasound-based technique in which the distribution of strain is mapped when soft tissue is exposed to repetitive stress. The method was described by Ophir et al. and referred to as quasi-static elastography [2]. Several methods for registering strain or elasticity using different forces to induce tissue deformation leading to recordable displacements or shear waves have subsequently emerged [3, 4]. These techniques have been used to characterize tissue in focal lesions, contracting organs, such as the heart [5, 6], stomach [7], and pancreas [8], or in liver fibrosis [9, 10]. Hard tissue is characterized by low strain, while softer areas show higher strain. RTE has shown promising results for differentiating between malignant and benign lesions in breast lumps [11], prostate tumors [12, 13], thyroid nodules [14–16], lymph nodes [17, 18] and the pancreas [8, 19]. Both malignant tumors and chronic inflammation may lead to the formation of soft tissue fibrosis [20, 21]. Fibrosis formation in malignant tumors is called desmoplasia and may contribute to increased tissue hardness through several mechanisms involving tumor cells, stromal cells and the extracellular matrix [22]. Malignant tumors are also usually characterized by increased interstitial pressure, a feature that decreases quickly in devascularized tissue [23–25]. In this study, RTE was performed on surgically resected bowel specimens containing benign or malignant lesions using one categorical classification system and two semi-quantitative methods for relative strain assessment. Our aim was to evaluate whether RTE could differentiate lesions caused by inflammation from malignant neoplastic lesions using qualitative and semi-quantitative methods for strain assessment. Furthermore, we examined whether strain ratio (SR) measurements were affected by changes in the elasticity dynamic range (E-dyn). Finally, we also wanted to correlate elastography results to a histological semi-quantification of fibrosis, inflammation parameters and tumor description.

Material and methods

Patients referred for elective bowel resection were recruited for this feasibility study from May 2008 to June 2009. All patients signed an informed consent to participate. The study was approved by the Regional Committee for Medical and Health Research Ethics in Western Norway and was conducted according to the Helsinki Declaration. 38 lesions from 27 patients were examined, 16 sections from Crohn’s stenoses in 9 different patients, 18 sections from adenocarcinomas in 16 patients and 4 sections from adenomas in 3 different patients. One cancer patient had a synchronous adenoma and another had two synchronous adenocarcinomas. Two adenomas were separate lesions from the same patient. In one patient with a multilobular tumor, two different tumor sections were included as separate lesions. In the Crohn’s disease specimens, separate skip lesions as well as the proximal and distal end of some longer stenoses were included as individual lesions. The observer was not blinded to the indication for surgery, but the final diagnosis was confirmed by histopathology after the examination and image evaluation.

Ultrasound equipment

A Hitachi Hi Vision 900 ultrasound scanner with software version V16–04 STEP 2 and a L54 M linear probe with frequencies 9–13 MHz (Hitachi Medical Corporation, Tokyo, Japan) were used in the study. The elastography modality was developed for free-hand application, and small repetitive compression/decompression movements of the probe were used to provide stress and induce tissue strain. With this method, relative strain values in different areas are calculated based on the assumption that stress is evenly distributed over the region of interest (ROI). A color map showing the strain in individual image elements as compared to the mean strain of the whole ROI is produced. The color distribution depends on the selected level of the E-dyn. A graphic indicator on the display shows when there is sufficient strain to produce an elastogram. The applied strain imaging algorithm records changes in echo positions between consecutive radio frequency frames using the extended combined autocorrelation method (ECAM), which is an extension of the original quasi-static elastography [26, 27]. We used the following parameter settings for RTE: Frame reject: 6, noise reject: 4, persistence: 2–5, smoothing: 2–3. The VAS and categorical scores were determined using E-dyn level 4 (default level). SR examinations were also recorded at E-dyn levels 2 and 3. The VAS and categorical scores were determined using E-dyn level 4 (default level). SR examinations were also recorded at E-dyn levels 2 and...
The third strain evaluation method was a semi-quantitative measurement of SR comparing the reference tissue (B) and the lesion (A) using the following formula:

\[
\text{Strain ratio} = \frac{\text{Mean strain in reference (B)}}{\text{Mean strain in lesion (A)}}
\]

SR is a built-in scanner software function and was first introduced in breast scanning and was called the “fat-lesion ratio”. It compares the mean strain in a user-selected portion of the reference tissue with a similarly selected part of the lesion, assuming that both areas have been subjected to the same amount of stress. We selected circular areas of reference tissue at the same depth and with approximately the same size as the lesion for all measurements (Fig. 2). In most cases, the central area of the lesion was selected for calculation (area A). However, if the lesions had reproducible differences in tissue hardness, the hardest area was selected as area A. Pericolic or mesenteric fat and connective tissue as well as parts of the normal bowel wall could be included in the reference area. SR measurements were performed on representative image frames with sufficient signal quality. Measurements were usually repeated 3 times on E-dyn levels 2, 4 and 6, respectively, thus yielding 9 SR measurements per lesion. However, only 3 SR measurements were recorded in three lesions. (E-dyn scale: 1 – 8).

**Methods for strain assessment**

We applied a categorical classification based on the color distribution within the lesion as originally described for pancreatic lesions by Janssen et al. [28]. In this classification, three basic patterns of color distribution are recognized: Type I is relatively homogeneous, type 2 has 2 or 3 different colors and type 3 has a honeycomb pattern. The colors present are represented by the subsequent letters usually indicating a declining proportion in the lesion: A = blue, B = green/yellow and C = red. 32 lesions (92 %) were classified according to this system and fell into five categories: 1B: homogeneously green, 2BA: green > blue, 2AB: blue > green, 1A: homogeneously blue and 2 A-B-C: presence of blue, green and red without one of the colors being classified as more prominent.

A visual analog scale (VAS) consisted of a 100-mm line where the observer marked the assumed level of the lesion’s hardness as assessed by elastography at E-dyn level 4. If the lesion was regarded as equal to the surrounding tissue, the mark was set at 50 mm. Lesions appearing harder were marked between 50 and 100 mm and softer lesions were marked between 0 and 50 mm. Assessments using the VAS score and visual categorical score were repeated by a trained observer blinded to the clinical data using video clips of the elastography seen in Windows Media Player (Microsoft Inc., Ca, USA). The results were compared to those obtained in the original scoring by the non-blinded observer.

**Statistical analysis**

The VAS scores were found to be within the limits of normal distribution for adenomas, Crohn’s lesions and adenocarcinomas as tested with the Shapiro-Wilk test, and comparisons between the groups were based on mean values with standard deviations in each group. When evaluating SR differences, the median of repeated SR measurements in each lesion was selected for further quantitative analysis. When comparing the three diagnostic groups of lesions, the mean of these SR medians was calculated within each group. The median SR results from Crohn’s lesions and adenocarcinomas were not distributed normally. A log transformation was attempted, but did not yield a normal distribution. We had included one to three separate lesions from the surgical specimens due to skip lesions in Crohn’s lesions and some synchronous neoplastic lesions. Consequently, the untransformed data were analyzed using a generalized estimating equations
(GEE) model allowing analysis of both related and unrelated observations. In this model, SR and VAS results for Crohn’s lesions and adenomas were compared to the corresponding results for adenocarcinomas, which served as a reference. For comparison of SR measurements at E-dyn levels 2, 4 and 6, we used a Spearman’s rho correlation and Wilcoxon-Mann-Whitney test for paired differences between the medians of three recorded SR measurements at each E-dyn level. The significance level was adjusted to apply for 3 measurements: \( p = \frac{0.05}{3} = 0.017 \). This test also provided information about the repeatability of SR measurements in the individual lesions.

The interobserver agreement for categorical scores was assessed using Cohen’s Kappa, which expresses the agreement controlled for agreement by chance and is considered a conservative measurement of interobserver agreement. The Kappa values were interpreted in the following manner: 0 = no agreement, 0 – 0.20 = slight agreement, 0.21 – 0.40 = fair agreement, 0.41 – 0.60 = moderate agreement, 0.61 – 0.80 = substantial agreement, and 0.81 – 1 = almost perfect agreement [29]. Continuous scores (VAS) were compared using correlation (Pearson’s) and limits of agreement [30]. Statistical analyses were performed using SPSS 17 and IBM-SPSS 19 (Chicago, Ill, USA).

**Results**

**Categorical evaluation of strain images**

The qualitative classification of color distribution in the elastogram showed reduced strain in Crohn’s lesions as well as adenocarcinomas indicating increased tissue hardness compared to surrounding tissue. Elastograms of an adenoma, an adenocarcinoma and a Crohn’s lesion are shown in Fig. 2a-c. Fig. 3 shows the distribution of qualitative strain image categories within each group of lesions evaluated by observer 1. 29 of 33 scored cases were described as category 2AB or 2 BA.
Semi-quantitative evaluation of strain images

The VAS evaluation showed no significant difference between Crohn’s lesions and adenocarcinomas, while adenomas had a significantly lower VAS score indicating softer tissue. SR quantification also could not demonstrate any difference between Crohn’s lesions and adenocarcinomas. However, the four adenomas had a significantly lower SR than the adenocarcinomas. The mean SR as well as median SR, intra-assay standard deviation, and VAS score for each diagnostic group are summarized in Table 1 and the GEE model details are listed for VAS and SR in appendices 1 and 2, respectively.

Interobserver validation of categorical and semi-quantitative elastography

For the categorical evaluation, cine loops of 28 of the original 33 lesions were re-evaluated by an observer blinded to the clinical data. Five cases were excluded from re-evaluation because only still images were stored and not the original video files. A comparison between observer 1 (original) and observer 2 (blinded) yielded a Kappa value of 0.38 for the categorical scale, indicating fair interobserver agreement. The second observer also repeated the VAS scoring in 30 of the 35 cases that had previously been scored by observer 1. There was a significant correlation between the two observers (Pearson’s r = 0.55, p = 0.002). The correlation plot and limits of agreement between observer 1 and 2 are shown in Fig. 4. Moderately good correlations were found with better interobserver agreement and correlation for higher values of VAS than for mid-range values.

Intraobserver validation of strain ratio measurements

The comparison between triplets of SR recorded in separate loops at different levels of E-dyn is presented in Table 2. We found moderate to good correlations (Spearman’s rho = 0.47 – 0.82) between the separate recordings with different E-dyn ranges. A comparison of the median SR showed no significant differences between recorded SR measurements between E-dyn levels 2 and 6 or between E-dyn levels 4 and 6. However, there was a significant difference in SR at E-dyn levels 2 and 4. The intra-assay standard deviations for all SR measurements in individual lesions are listed according to diagnosis in Table 1. These values showed higher measurement variation for Crohn’s lesions than for adenocarcinomas.

Assessment of fibrosis

The number of included patients, scanned sections, the macroscopic length and the categorical fibrosis score for each of the three diagnostic groups are presented in Table 3. There was no significant difference in the macroscopic length of Crohn’s steno-
ses and the maximum length of the adenocarcinomas. The adenomas were smaller than the adenocarcinomas and Crohn’s lesions. In two separate evaluations, one pathologist visually categorized the fibrosis in 37 sections of the same lesions as “moderate fibrosis” in 25 and 17 cases, respectively. Category 1 = “slight fibrosis” was used 7 and 13 times, respectively. The two separate evaluations with categorical fibrosis scores 0 – 3 had a Kappa value of 0.37. In the Crohn’s lesions, the fibrosis score indicated variations in fibrosis in different wall layers. The mean

![Fig. 4](Correlation plot and limits of agreement plot of VAS score agreement between observer 1 and observer 2. Observer 2 was blinded to the clinical data. The agreement improved with higher VAS scores.

**Abb. 4** Korrelationsplot und „Limits of agreement“-Plot der Übereinstimmung von Beobachter 1 und 2 bei der VAS-Auswertung. Beobachter 2 kannte die klinischen Daten nicht (Verblindung). Die Übereinstimmung verbesserte sich mit höherer VAS-Punktzahl.

<table>
<thead>
<tr>
<th>Diagnosis/Adenocarcinoma</th>
<th>Patients</th>
<th>Lesions</th>
<th>Macroscopic Length (cm)</th>
<th>Fibrosis Score (0–3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Score 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Score 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Score 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Score 3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diagnosis/Crohn’s Lesion</th>
<th>Patients</th>
<th>Lesions</th>
<th>Macroscopic Length (cm)</th>
<th>Fibrosis Score (0–3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Score 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Score 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Score 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Score 3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diagnosis/Adenoma</th>
<th>Patients</th>
<th>Lesions</th>
<th>Macroscopic Length (cm)</th>
<th>Fibrosis Score (0–3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Score 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Score 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Score 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Score 3</td>
</tr>
</tbody>
</table>

**Table 2** Comparison of median strain ratio measurements of the same lesions with elasticity dynamic range (E-dyn) levels 2, 4 and 6. The differences were not normally distributed. The exact two-tailed Wilcoxon-Mann-Whitney test was used for the comparison of paired differences. The significance level for differences was set to p ≤ 0.05/3 = 0.017 to adjust for multiple comparisons.

<table>
<thead>
<tr>
<th>E-dyn levels compared</th>
<th>Pairs</th>
<th>Median SR (min-max)</th>
<th>Spearman’s correlation: rho (p-value)</th>
<th>P-value of paired differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-dyn 4</td>
<td>34</td>
<td>2.25 (0.68 – 8.28)</td>
<td>0.47 (0.005)</td>
<td>0.008</td>
</tr>
<tr>
<td>E-dyn 2</td>
<td>35</td>
<td>2.25 (0.68 – 8.28)</td>
<td>0.55 (0.001)</td>
<td>0.849 (n.s)</td>
</tr>
<tr>
<td>E-dyn 6</td>
<td>34</td>
<td>2.06 (0.33 – 22.75)</td>
<td>0.82 (&lt;0.001)</td>
<td>0.022 (n.s)</td>
</tr>
</tbody>
</table>

**Table 3** Descriptive data for the diagnostic entities adenocarcinoma, Crohn’s lesion and adenoma including the number of patients, number of lesions, macroscopic length and number of lesions categorized using the 0 – 3 fibrosis score by the pathologist. One patient had both an adenocarcinoma and an adenoma and was counted twice. A total of 27 patients were included.

**Table 3** Descriptive data for the diagnostic entities adenocarcinoma, Crohn’s lesion and adenoma including the number of patients, number of lesions, macroscopic length and number of lesions categorized using the 0 – 3 fibrosis score by the pathologist. One patient had both an adenocarcinoma and an adenoma and was counted twice. A total of 27 patients were included.

Havre RF et al. Strain Assessment in... Ultraschall in Med 2014; 35: 149–158
fibrosis scores were as follows: submucosa – 2.44, subserosa – 1.75 and proper muscle – 1.50.

Assessment of inflammation in Crohn’s specimens

The patients diagnosed with Crohn’s disease all had pain as the main symptom with stenotic findings in the small or large bowel confirmed by radiological imaging or endoscopy. Eight of nine patients were non-responders to high-dose steroids or TNF-antagonists. Two of the patients who underwent surgery in 2008 were reluctant to use TNF-antagonists or prednisolone. Two patients who underwent surgery only 3 and 8 months after diagnosis had suffered substantial weight loss of 10 and 15 kg, respectively. One Crohn’s patient was operated on under the suspicion of cancer in the ascending colon, although histopathology had not confirmed this prior to surgery. This patient had only received a 10-day course of antibiotics prior to surgery. All Crohn’s lesions exhibited transmural inflammatory changes. Lymphoplasmacytic inflammation was assessed as slight in 4/16 cases, moderate in 8/16 cases and severe in 4/16 cases. Other inflammatory parameters were mucosal ulcerations in 13/16 (81%) cases, granulomatous tissue in 12/16 (75%) cases, cryptitis in 7/16 (44%) cases, fissure ulcers in 4/16 (25%) cases, fistulas in 3/16 (19%) cases and abscesses in 0/16 cases. Lesions with high overall inflammatory activity did not differ from lesions with low overall inflammatory activity by fibrosis score. Furthermore, there was no difference in SR or VAS between lesions with overall high or low inflammatory activity. Inflammatory changes in the adenocarcinomas and adenomas were not specifically evaluated. However, none of these specimens exhibited transmural chronic inflammatory changes. Only one adenocarcinoma specimen also contained an area of acute inflammation and abscess formation according to the original pathology report.

Strain evaluation by tumor stage and grade

The SR and VAS results were compared by tumor stage according to the TNM classification for the 18 examined sections of malignant lesions. The results are summarized in Table 4. There was a tendency for a higher fibrosis score for tumors staged as T3 or T4. We found no significant differences in SR and VAS between lesions staged as T3 or T4 compared to T1 and T2, but the included numbers were low. The tumor grade was described as highly differentiated (>95% glandular structures) in 2 cases, moderately differentiated (50–95% glandular structures) in 9 cases and poorly differentiated (<5% glandular structures) in 7 cases. When highly and moderately differentiated tumors were considered as one group (n = 11) and compared to poorly differentiated tumors (n = 7), we found no significant differences in the SR or VAS scores. The number of lesions given fibrosis score 2 was 8/11 (73%) for the highly and moderately differentiated adenocarcinomas and 5/7 (71%) for the poorly differentiated adenocarcinomas.

### Table 4

<table>
<thead>
<tr>
<th>T-stage</th>
<th>n</th>
<th>fibrosis score = 1</th>
<th>fibrosis score = 2</th>
<th>median SR (min-max)</th>
<th>VAS score (mean, SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1.73 (1.62–1.84)</td>
<td>62.5 (4.95)</td>
</tr>
<tr>
<td>T2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1.56 (0.81–2.31)</td>
<td>72.0 (16.97)</td>
</tr>
<tr>
<td>T3</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>3.00 (1.19–5.34)</td>
<td>81.1 (11.05)</td>
</tr>
<tr>
<td>T4</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>2.41 (1.59–3.18)</td>
<td>72.2 (15.29)</td>
</tr>
</tbody>
</table>

Discussion

Differentiation between neoplastic and inflammatory lesions by RTE

RTE demonstrated low strain indicating harder tissue in resected lesions from both colorectal carcinomas and Crohn’s disease. The hypothesis that strain imaging could differentiate malignant neoplastic lesions from benign lesions caused by inflammation was based upon observations that malignant tumors in other organs tended to become harder than benign lesions [14, 19, 31–33]. Pathologists frequently observe increased fibrosis within and in the vicinity of malignant lesions. A similar process also occurs in chronic inflammation in which case fibroblasts and possibly other cells produce collagen-rich interstitium resulting in scar tissue [34]. Such fibrogenic activity seems to be a dominant feature in stenotic lesions of Crohn’s disease. A second contributor to the hard appearance of malignant tumors may be high interstitial pressure [23, 25, 35, 36], which decreases quickly after devascularization. Hence, its probable contribution to differences in tissue elasticity was not measureable in our study on ex-vivo tissue specimens. Crohn’s disease is usually subjected to surgery because of intractable stenosis not responding to medical therapy. All Crohn’s lesions in our study had extensive fibrosis, as demonstrated by histology. Hence, increased tissue hardness was visualized in both Crohn’s lesions and adenocarcinomas. However, it was not possible to detect any strain differences between benign transmural lesions caused by chronic inflammation and malignant tumors.

Strain evaluation and fibrosis in adenocarcinomas

The number of adenocarcinomas in the study was limited. Hence, we could not find a statistically significant difference in SR or VAS score, but there was a tendency to increase the fibrosis score and SR with increasing tumor stage. With respect to tumor grading, we would expect poorly differentiated tumors to have a higher fibrosis score. However, no significant strain differences were found, even if both the SR and VAS scores showed an insignificant tendency towards higher strain for poorly differentiated tumors. The low number of lesions included may be the cause for this (type 2 error), and larger studies are warranted.

Strain assessment

A majority of published papers on RTE have used visual categorisation based on color distribution in the lesion as the main strain evaluation tool. In this study, we applied a categorical classification based on a distribution pattern and predominance of separate colors within the lesions. This classification was introduced by Janssen et al. [28] for the evaluation of pancreatic lesions using endoscopic ultrasonography, and up to five of the original categories were used in our study. The qualitative scoring system could not reliably separate the diagnostic entities in this study. Observer 1 classified 50% of the adenomas, 50% of the Crohn’s lesions and 35% of the adenocarcinomas as 2BA, indicating more green than blue. Category 1B, with a homogeneously

Havre RF et al. Strain Assessment in... Ultraschall in Med 2014; 35: 149–158
green presentation, was only used for benign lesions, but only 3 of 16 benign lesions were categorized as 1B. The interobserver reproducibility of categorical scoring was fair (Kappa: 0.38) when the cine loops were re-evaluated by a second observer. 

Elastograms were also evaluated with a continuous VAS score, anticipated to be more fine-grained than the categorical score, and comprising an aspect of strain comparison between the lesion and the surrounding reference tissue. Because none of the lesions were perceived as softer than the reference tissue, the lower half of the scale was not utilized. VAS evaluation had a moderate correlation (r = 0.55) between the original and an observer blinded to clinical data. The difference between the observers was larger for low scores (moderately hard and mixed lesions) than for high scores (hardest lesions) (Fig. 4). This coincides with an observation that two observers frequently scored the same lesions as 2AB (inhomogeneous, blue > green) and 2BA (inhomogeneous, green > blue), correspondingly, on the categorical scale.

The SR is based on a computer-based comparison of mean strain in two selected areas, and may thus seem to be more objective than the VAS and the categorical scale. However, the selection of calculation areas is still user-dependent, and the recorded strain may be subject to significant variations especially in heterogeneous reference areas. Uneven application of stress within the elastogram ROI may also influence the SR results. Hence, SR evaluation is usually quoted as a semi-quantitative method. The SR data (Table 1) demonstrates a large spread in strain measurements within the groups of Crohn’s lesions and adenocarcinomas. The variation between cases in the same group may express measurement variation and also reflect a biological diversity in tissue strain for that entity.

Reproducibility of SR measurements

The E-dyn level of RTE influences the color presentation of strain distribution profoundly [37]. Consequently, it is crucial to keep this setting on a standardized level for the visual comparison of strain images in different cases. Median SR recordings at E-dyn 4 vs. 6 and E-dyn 2 vs. 6 were not significantly different. A significant difference was found for the SR recorded at E-dyn 2 vs. 4. This was an unexpected and probably unreliable result since such a difference was not found with a larger E-dyn span between 2 and 6. We have previously shown that SR measurements are independent of E-dyn level in a standardized liver-tissue-mimicking phantom [38]. Therefore, we interpreted the observed SR measurement variability as the result of the general variability of this method in separate recordings of the same lesions using a free-hand application, and not as a consequence of the selection of different levels of E-dyn. More dedicated studies may be required to confirm the relationship between E-dyn level and SR measurements in biological tissues.

Relevance

The differentiation between colorectal tumors and Crohn’s lesions predominantly in the small intestines may seem to have limited clinical value. However, these lesions were chosen as a model for lesions caused by benign inflammation and malignant neoplasms in the bowel wall. We selected Crohn’s lesions because they are transmural and expand the bowel wall in a similar manner to neoplasias. Rustemovic et al. have studied patients with ulcerative colitis and compared them to patients with Crohn’s disease and healthy controls in vivo [39]. They found a significantly higher SR in Crohn’s patients than in ulcerative colitis patients with active rectal disease. In the case of Crohn’s disease, the patients with active disease also had a significantly higher SR than those in remission, indicating a substantial contribution from increased interstitial pressure. In the case of ulcerative colitis, the inflammation is limited to the mucosa and submucosa and does not create the transmural inflammation and fibrosis characteristic for Crohn’s disease.

From our point of view, the contribution of endoscopic RTE in the differentiation between Crohn’s disease and ulcerative colitis may not provide more than visualization of the transmural inflammation resulting in bowel wall thickening clearly visible by B-mode EUS. However, further studies may investigate whether a possible clinical application of EUS-elastography may be the evaluation of polyps and pseudo-polyps in ulcerative colitis, and whether a thickened bowel wall with increased tissue hardness may indicate the development of malignancy in a patient with ulcerative colitis under surveillance.

Very few adenomas were included in this study, as the main object was to evaluate Crohn’s lesions and adenocarcinomas. However, we found that adenomas appeared to have higher strain and, correspondingly, lower strain ratios than both adenocarcinomas and Crohn’s lesions. The difference in strain properties between adenomas and adenocarcinomas in the rectum is clinically relevant, and endorectal ultrasonography (ERUS) with elastography can potentially be a valuable preoperative diagnostic tool [33]. ERUS is currently acknowledged as the best method for selecting patients for minimally invasive rectal surgery such as endoscopic mucosal resection or transanal endoscopic microsurgery [40, 41]. The preoperative classification of early rectal adenocarcinomas and adenomas remains challenging and there is still a need for improved methods in the staging of these tumors.

Limitations of the study

This study was performed on surgically resected specimens, in which the contribution of interstitial pressure as a source of tissue hardness is eliminated. Hence, the ex-vivo setup may underestimate tissue hardness in vascularized tissue. However, all surgical specimens were examined under the same conditions, and the contribution from interstitial pressure should be reduced in a similar way for all samples. A second limitation is that the primary observer was not blinded to the clinical diagnosis. A second observer, blinded to clinical data, re-evaluated the elastography and B-mode video loops. The second evaluation was performed during a short time span (days) while the original scoring was performed after each examination spanning a period of 12 months. This may have caused some drifting in visual scoring over time, and may partly explain the moderate agreement for VAS and the fair agreement for the categorical score. The low number of adenomas included in this study (4) limits the validity of the statistical difference found between adenomas and adenocarcinomas.

Conclusion

Free-hand RTE of newly resected surgical specimens could not differentiate stenotic lesions caused by chronic inflammation in Crohn’s disease from adenocarcinomas on the basis of differences in semi-quantitative strain measurements. The SR and VAS score showed significant strain differences between adenocarcinomas and a limited number of adenomas. Pathological evaluation showed increased tissue fibrosis in both adenocarcinomas and Crohn’s lesions. The interobserver reproducibility of the visual evaluation using a VAS score and categorical scale was fair to very good.
moderate. The intra-assay reproducibility of SR was more reliable, but the potential influence of E-dyn settings on SR measurements could not be unequivocally affirmed. No significant differences in strain assessment were found on the basis of inflammation grade in Crohn’s lesions and by tumor stage or grade in adenocarcinomas. Ultrasound elastography is a promising clinical method, but increased strain may be observed in malignant and benign lesions and further studies on strain data of different lesions in vivo are warranted.

Abbreviations

RTE real-time elastography
SR strain ratio
E-dyn elasticity dynamic range
VAS visual analog scale

Appendix

### Appendix 1

<table>
<thead>
<tr>
<th>tissue strain by visual analog scale (VAS), n = 35 (GEE, linear model)</th>
<th>test of model effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>diagnostic group</td>
<td>estimated marginal means</td>
</tr>
<tr>
<td>adenoma (n = 4)</td>
<td>59.8</td>
</tr>
<tr>
<td>Crohn’s lesion (n = 13)</td>
<td>81.6</td>
</tr>
<tr>
<td>adenocarcinoma (n = 18)</td>
<td>75.2</td>
</tr>
</tbody>
</table>

### Appendix 2

<table>
<thead>
<tr>
<th>tissue strain measured by strain ratio (SR), n = 38 (GEE, linear model)</th>
<th>test of model effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>diagnostic group</td>
<td>estimated marginal means</td>
</tr>
<tr>
<td>adenoma (n = 4)</td>
<td>1.31</td>
</tr>
<tr>
<td>Crohn’s lesion (n = 16)</td>
<td>3.33</td>
</tr>
<tr>
<td>adenocarcinoma (n = 18)</td>
<td>2.51</td>
</tr>
</tbody>
</table>

Affiliations

1 Institute of Medicine, University of Bergen
2 The Gade Institute, University of Bergen
3 National Centre for Ultrasound in Gastroenterology, Department of Medicine, Haukeland University Hospital, Institute of Medicine, Bergen
4 Department of Surgery, Haukeland University Hospital, Bergen
5 Institute of Surgery, University of Bergen

Acknowledgement

We thank the staff at Department of Abdominal and Emergency Surgery, Surgical Clinic, Haukeland University Hospital for their cooperation in recruiting patients and securing resected specimens. We also appreciate valuable statistical counseling from Geir Egil Eide at Centre for Clinical Research, Haukeland University Hospital.

References

29 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159–174