Ann Surg Oncol (2019) 26:1795–1804
https://doi.org/10.1245/s10434-019-07292-0

Annals of
SURGICAL ONCOLOGY
OFFICIAL JOURNAL OF THE SOCIETY OF SURGICAL ONCOLOGY

ORIGINAL ARTICLE – GASTROINTESTINAL ONCOLOGY

Can We Accurately Identify Peritoneal Metastases Based on Their Appearance? An Assessment of the Current Practice of Intraoperative Gastrointestinal Cancer Staging

Thomas Schnelldorfer, MD, PhD[1,2], Matthew P. Ware, BS[1], Li Ping Liu, PhD[3], Michael G. Sarr, MD[4], Desmond H. Birkett, MD[1], and Robin Ruthazer, MPH[5]

[1]Department of Surgery, Lahey Hospital and Medical Center, Burlington, MA; [2]Department of Biomedical Engineering, Tufts University, Medford, MA; [3]Department of Computer Science, Tufts University, Medford, MA; [4]Department of Surgery, Mayo Clinic, Rochester, MN; [5]Clinical and Translational Science Institute, Tufts University, Boston, MA

ABSTRACT

Background. Peritoneal lesions are common findings during operative abdominal cancer staging. The decision to perform biopsy is made subjectively by the surgeon, a practice the authors hypothesized to be imprecise. This study aimed to describe optical characteristics differentiating benign peritoneal lesions from peritoneal metastases.
Methods. The study evaluated laparoscopic images of 87 consecutive peritoneal lesions biopsied during staging laparoscopies for gastrointestinal malignancies from 2014 to 2017. A blinded survey assessing these lesions was completed by 10 oncologic surgeons. Three senior investigators categorized optical features of the lesions. Computer-aided digital image processing and machine learning was used to classify the lesions.
Results. Of the 87 lesions, 28 (32%) were metastases. On expert survey, surgeons on the average misidentified 36 ± 19% of metastases. Multivariate analysis identified degree of nodularity, border transition, and degree of transparency as independent predictors of metastases (each $p < 0.03$), with an area under the receiver operating characteristics curve (AUC) of 0.82 (95% confidence interval [CI], 0.72–0.91). Image processing demonstrated no difference using image color segmentation, but showed a difference in gradient magnitude between benign and metastatic lesions (AUC, 0.66; 95% CI 0.54–0.78; $p = 0.02$). Machine learning using a neural network with a tenfold cross-validation obtained an AUC of only 0.47.
Conclusions. To date, neither experienced oncologic surgeons nor computerized image analysis can differentiate peritoneal metastases from benign peritoneal lesions with an accuracy that is clinically acceptable. Although certain features correlate with the presence of metastases, a substantial overlap in optical appearance exists between benign and metastatic peritoneal lesions. Therefore, this study suggested the need to perform biopsy for all peritoneal lesions during operative staging, or at least to lower the threshold significantly.

Operative staging with correct identification of peritoneal metastases is crucial in the selection of appropriate treatment for patients with potentially resectable abdominal malignancies. Because visually evident lesions or abnormalities of the peritoneal lining are a common finding, even in the absence of cancer, the surgeon typically is left with a subjective decision when to perform biopsy of what is believed to be a suspicious-appearing lesion and when to consciously ignore what is presumed to be a benign-appearing lesion.

Historically, these determinations have been made subjectively based on features of the lesion's gross visual appearance that are presumed to allow categorization into benign, questionable, or malignant lesions despite the absence of true scientific evidence supporting this approach. This presumption has led to a widely accepted approach in line with the famous 1964 quote by Justice Stewart Potter of the United States Supreme Court: "I know it when I see it." In reality, however, for patients with a diagnosis of pancreatic, gastric, or gallbladder

First Received: 2 October 2018;
Published Online: 25 March 2019

T. Schnelldorfer, MD, PhD
e-mail: thomas.schnelldorfer@lahey.org

carcinomas, for example, this approach is evidenced to be crude and imperfect, especially considering that metastatic progression is historically detected on cross-sectional imaging in up to 40% of these patients within 12 months after an R0 resection, with the peritoneum a common site for initial "recurrence."[1] In these cases, it can be assumed that most of these peritoneal metastases likely were present at the time of the initial operation.

Because identification of cancer metastases is an important part of appropriate cancer care, it is crucial that clinicians be able to identify which lesions are benign and thus can be ignored and which lesions have a probability of harboring malignancy and thus should be biopsied. This study aimed to determine the optical appearance specific to peritoneal metastases and the accuracy of using the gross appearance of a peritoneal lesion to determine the need for biopsy. The study was performed under the hypothesis that a significant number of peritoneal metastases are missed during routine operative staging due in part to incorrect categorization of a presumed benign lesion.

METHODS

Study Eligibility Criteria

We identified all patients who underwent an operative intervention for a biopsy-proven malignancy originating from the gastrointestinal tract (excluding esophageal, hepatic, and colorectal malignancies) at Lahey Hospital and Medical Center between 1 January 2014 and 30 September 2017 using the cancer registry and electronic medical records. Of 264 patients identified, 71 had at least one peritoneal biopsy performed with the intent to exclude metastases at the time of operation. Of those 71 patients, 37 underwent a laparoscopic operation with the initial intent for either resection or palliation of the underlying malignancy.

For 35 of the 37 patients, video recordings of the operation were available, allowing for review of the staging laparoscopy within a study cohort of 35 patients. All these operations were performed by a single surgeon (T.S.) according to routine clinical care using the Olympus Endoeye laparoscope with an Evis Exera II imaging platform (Olympus, Center Valley, PA, USA).

For each study patient, the entire video recording of the operation was evaluated, and two representative still images were saved in PNG file format of each lesion, which was biopsied during the operation. The first image prioritized getting a clear view of the lesion. These images were used for the expert survey, evaluation of optical appearance, color analysis, and machine learning. The second image included the lesion and a nearby surgical instrument of known size. These images were used for objective estimation of size and gradient magnitude. The study was approved by the Lahey Clinic Institutional Review Board.

Expert Survey

The de-identified and unlabeled electronic study images of peritoneal lesions were shown in randomized order to gastrointestinal oncologic surgeons of various subspecialties practicing at Lahey Hospital and Medical Center and Mayo Clinic Rochester. Surgeons were recruited until the goal of 10 recruits was achieved. The average time in practice for these experienced surgeons was 27 years (range 6–45 years).

The 10 surgeons were aware that the study group consisted of patients with gastrointestinal malignancies, but they were blinded to the fact that all those lesions had been biopsied and to the pathologic results of the biopsy. The study images were presented to these surgeons in size comparable with the standard monitor setups during laparoscopy. The surgeons were prompted to the specific site of the lesion in question. They were asked to classify the lesion using a 5-point Likert-type scale, with categorization of each lesion as "extremely likely benign," "likely benign," "neutral," "likely metastasis," or "extremely likely metastasis." In addition, the surgeons were asked whether they would have biopsied the lesion.

For purposes of the analysis, the lesions categorized as "extremely likely benign," "likely benign," or "neutral" were grouped as benign, whereas the lesions categorized as "likely metastasis" or "extremely likely metastasis" were grouped as metastasis.

Subjective Evaluation of Optical Appearance

Three experienced investigators blinded to the pathologic results of the biopsy characterized the appearance of the imaged lesions. Discrepancies between the three evaluations were resolved by group consensus via re-review. The variables evaluated were those commonly perceived to be associated with metastases. Because the literature had no formal definition for any of these variables, categorization was determined subjectively in accordance with clinical practice.

Digital Image Processing

Measurement of Lesion Size Using ImageJ software version 1.50 g (National Institute of Health, Bethesda, MD, USA), the size of the lesion was obtained by measuring both the long axis (defined as the longest pixel length of each lesion) and the short axis (defined as the longest pixel length of each lesion perpendicular to the long axis). Using the pixel length of a nearby surgical instrument of known size as a reference, the actual size of each lesion was estimated. The surface area was estimated using the following formula: $\pi \times$ (Length of long axis/2) $\times$(Length of perpendicular axis/2).

Image Segmentation Color Analysis The PNG files were analyzed using ImageJ software, and the red, green, blue (RGB) images were converted to gray scale and HSV images. The RGB and HSV channels of the eight-bit graphics were separated, together with the gray-scale images, resulting in seven images representing the intensity values (0–255) of the combined color, red, green, and blue channels as well as hue, saturation, and value/brightness. The region of interest (ROI, i.e., the area of the visible lesion) and a rim of healthy tissue immediately adjacent to the lesion were marked, respectively. The intensity values of all the pixels within the marked areas of each image segment were obtained.

Measurement of Gradient Magnitude Changes in luminosity at the edge of each lesion were measured perpendicularly from the edge inward at a location representing the typical transition of the lesion's color using gray-scale images. The actual distance of the maximal change in luminosity was calculated using the pixel length of the detected change compared with the pixel length of a nearby surgical instrument of known size. The gradient was reported as the change in luminosity per actual distance.

Machine Learning

Using the RGB images, a rectangular bounding box was placed around the visible lesion. The image within each bounding box was used as a patch and resized to 128 × 128 pixels. A deep neural network (DNN) was trained to classify these patches. The DNN model had three convolution layers followed by two dense layers. Each convolution layer consisted of a convolution operation with rectified linear unit activation and a max-pooling operation. The first dense layer had an output size of 32, and the second dense layer predicted the class of the patch. The dropout rate after the first dense layer was 0.5. A tenfold cross-validation was run on the 87 training patches, and the results were reported. The model was selected by varying the structure and parameters of the network on the training set of the first fold. For the neural network, Keras, an open-source deep learning library, was used.[2]

Statistical Methods

The initial design of the study required enrollment of approximately 90 image samples, assuming one-third to be in the metastasis group and two-thirds in the benign group to detect an absolute difference of 30% for any given characteristic between the two groups with a power of 80% and an alpha level of 0.05. The objective and subjective characteristics of the lesions and patient characteristics are presented using medians with ranges, means and standard deviations, or percentages falling within a subgroup.

Data were summarized separately for pathologically confirmed benign and metastatic lesions. Raw comparison of lesion dimensions used the non-parametric Kruskal–Wallis test because of skewed distributions and was summarized as a median. Unadjusted and adjusted associations of lesion characteristics with type of lesion (metastatic vs benign) were estimated using generalized estimated equations (GEE) to account for multiple lesions coming from individual patients.

Multivariable GEE models for the outcome of metastatic (vs benign) lesions were constructed using a backward selection process that began by including variables with a p value lower than 0.1 for unadjusted associations. Lesion dimensions were log transformed for the GEE analyses to normalize the skewed distributions. The area under the receiver operating characteristics curve (AUC) was calculated as a measure of the statistical models to discriminate between metastatic and benign lesions. The Youden index was used to identify cut points for optimal sensitivity and specificity of a test based on predictions from the statistical models.

Mean pixel intensity values were compared between benign lesions and metastases using both Student's t tests and GEE models to account for multiple lesions per person. Kappa statistics together with 95% confidence intervals (CIs) were used to quantify agreement between raters. All statistical analyses were performed using SAS Enterprise Guide version 7.15 software (SAS Institute Inc, Cary, NC, USA) and Microsoft Excel for Mac version 14.4.0 (Microsoft, Redmond, WA, USA).

RESULTS

The demographics and clinical findings of the 35 study patients are described in Table 1. Biopsy of 88 peritoneal lesions in the study cohort was performed. The pathologic result of the peritoneal biopsy was indeterminate in one case, and this lesion was therefore excluded. Of the 87 remaining peritoneal lesions representing the study sample, 28 were metastases and 59 showed benign changes at the final pathology. Of the 59 benign lesions, 14 occurred in the setting of a patient with other biopsies showing peritoneal metastases. Regarding the origin of the lesion, 82% originated from the parietal peritoneum and 18% from the visceral peritoneum.

TABLE 1 Patient demographics and clinical findings

Median age: years (range)	67 (44–85)
Sex (male/female)	23/12
Primary neoplasm (no. of patients)	
Gastric adenocarcinoma	19
Pancreatic adenocarcinoma	11
Gallbladder carcinoma	2
Metastatic pancreatic neuroendocrine tumor, jejunal adenocarcinoma, ampullary adenocarcinoma	1 each
Median no. of biopsies performed per patient (range)	2 (1–5)
Peritoneal Cancer Index (no. of patients)	
No evidence of metastases	20
Score 1–5	10
Score 6–10	1
Score > 10	4

Human Assessment

Expert Survey The experts categorized lesions as "extremely likely benign" 12% of the time, "likely benign" 37% of the time, "neutral" 19% of the time, "likely metastasis" 22% of the time, and "extremely likely metastasis" 10% of the time. The results for the median rate of correct classification for each lesion are shown in Fig. 1. The grand median rate of correct classification was only 70%. Certain lesions were seemingly easier to classify than others, as demonstrated by a wide range of median rates between 0 and 100%. In addition, considerable variability existed among the experts, confirmed by only fair interobserver agreement (kappa, 0.35; 95% CI 0.32–0.39).[3]

Regarding the performance of each oncologic surgeon, the average rate of correct classification was only 64 ± 19% for metastases and only 82 ± 11% for benign lesions, resulting in an average accuracy of only 76 ± 6%. Hence, on the average, 36 ± 19% of metastases were not recognized.

The findings showed a trend toward surgeons who correctly identified metastases at a greater rate (i.e., greater sensitivity), having a greater rate of false-positive results (i.e., less specificity) and a seemingly lower threshold for calling a lesion suspicious and for biopsy of any lesion (Fig. 2), suggesting that improved identification of metastases might not be due to improved accuracy but rather to an issue of threshold. Surgeons seemingly recognized the imperfection of their clinical assessment, as demonstrated by the finding that 42 ± 15% of lesions not thought to be metastases still were recommended by the surgeon for biopsy. Regarding the decision to perform biopsy of lesion, the oncologic surgeons on the average would not have biopsied 16 ± 10% of metastases, but would have biopsied 42 ± 15% of benign lesions, leading to an accuracy of only 67 ± 8% and a clinically important number of metastases that would have been missed by their visual appearance alone.

Subjective Evaluation of Optical Appearance A total of 10 optical features were assessed. The results are outlined in Table 2. The following six evaluated features were identified as associated with the presence of metastases at a p value level lower than 0.1: degree of nodularity, predominant degree of transparency, transition in color at the border, color spectrum, predominant spatial relationship to microvasculature, and neovascularization. For these six variables, the sensitivity and thus the rate of the specific optical feature present in peritoneal metastases ranged only between 61 and 93%. The positive predictive value and thus the probability of the specific optical feature

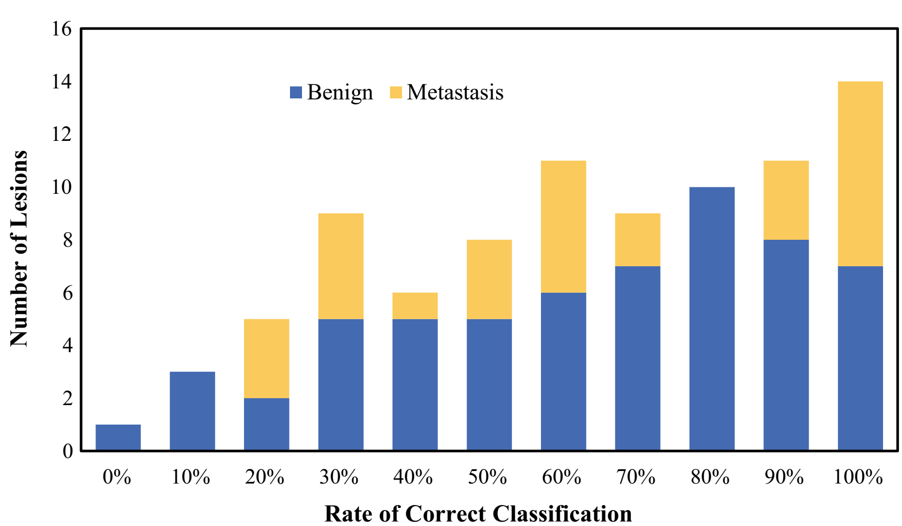


FIG. 1 Expert survey results showing the distribution of the mean rate for correct classification of each of the 87 lesions' (i.e., the figure stratifies the cohort according to the rate of correct classification showing the number of lesions within each stratum). The findings demonstrate that some lesions were more difficult to classify than others (i.e., some lesions were correctly classified by the majority of experts, whereas others were not)

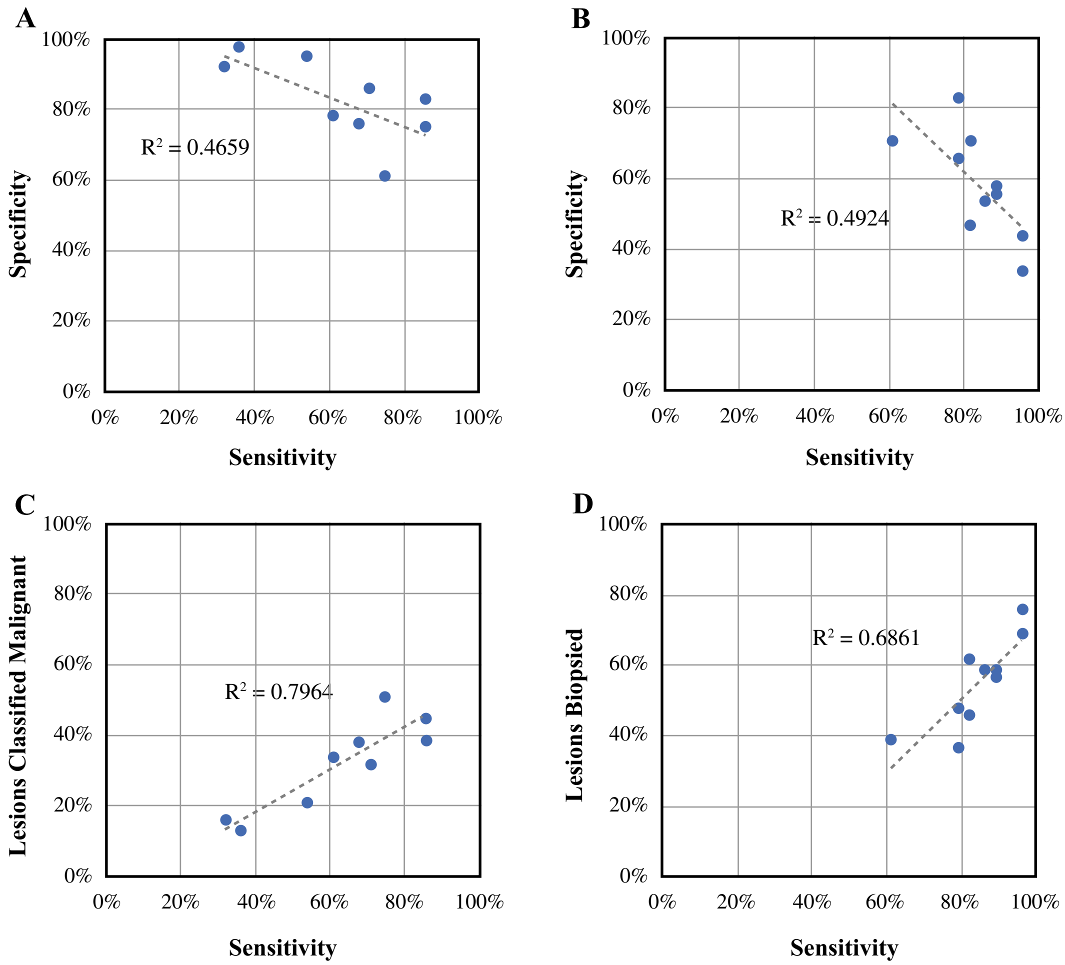


FIG. 2 Expert survey results showing the performance of 10 surgeons evaluating 87 peritoneal lesions. Each data point represents the performance of a single surgeon. **a** Comparison between the rate for correct classification of metastases (sensitivity) and the rate for the correct classification of benign lesions (specificity). **b** Comparison between the rate of suggested biopsy for metastases (sensitivity) and the rate of suggested non-biopsy for benign lesions (specificity). **c** Comparison between the rate for correct classification of metastases (sensitivity) and the fraction of all lesions labeled malignant. **d** Comparison between the rate of suggested biopsy for metastases (sensitivity) and the fraction of all lesions recommended for biopsy. The results demonstrate a correlation between increased detection of metastases with a lower threshold for classifying a lesion as metastasis (**a, c**) and a lower threshold for biopsy of a lesion (**b, d**)

predicting the presence of a peritoneal metastasis in these six variables ranged only between 10 and 83%. Similarly, the accuracy was only as high as 72%.

These six variables were included in the first step of a backward elimination process to create a multivariate model. The final multivariable model based on a minimum QIC statistic (quasilikelihood under the Independence Model Criterion) included three variables that were independent predictors for the presence of metastases: degree of nodularity (odds ratio [OR], 3.10; 95% CI 1.50–6.38; $p = 0.002$), transition in color at the border (OR, 4.78; 95% CI 1.62–14.11; $p = 0.005$), and predominant degree of transparency (OR, 4.89; 95% CI 1.26–19.00; $p = 0.022$). The AUC of this multivariable model was 0.82 (95% CI 0.72–0.91).

A second model was created for potential clinical use establishing a risk score ranging from 0 to 4, determined by the sum of points derived from the variables identified in the multivariable model (Table 3). Using only the point score as a single variable in a model, the odds ratio for an increase of 1 point was 3.62 (95% CI 1.89–6.93; $p < 0.0001$). Although the risk score was strongly associated with the presence of metastases, the AUC of this model was only 0.80 (95% CI 0.71–0.90), suggesting that the model, although ostensibly the best currently available predictive method, is only a fair classifier. For example, using only the highest risk features (i.e., score of 4 denoting a markedly nodular lesion, gradual color transition at border, and opaque/nontranslucent lesion), the accuracy would be only 78%, with a sensitivity under this criterion

TABLE 2 Gross appearance of peritoneal lesion seen by standard laparoscopy[a]

	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
Dimension					
Predominant shape ($p = 0.267$)					
Round or oval ($n = 72$)	89	20	35	80	43
Elongated ($n = 15$)	11	80	20	65	57
Symmetry ($p = 0.785$)					
Symmetric ($n = 36$)	39	58	31	67	52
Asymmetric ($n = 51$)	61	42	33	69	48
Predominant outline ($p = 0.584$)					
Smooth ($n = 34$)	32	58	26	64	49
Irregular borders ($n = 41$)	50	54	34	70	53
Spiculated ($n = 12$)	18	88	42	69	66
Degree of nodularity ($p = 0.002$)					
Flat ($n = 25$)	11	63	12	60	46
Slightly nodular ($n = 31$)	29	61	26	64	51
Markedly nodular ($n = 31$)	61	76	55	80	71
Surface texture ($p = 0.123$)					
Smooth ($n = 65$)	64	20	28	55	34
Irregular ($n = 22$)	36	80	45	72	66
Color					
Predominant degree of transparency ($p = 0.011$)					
Opaque ($n = 66$)	93	32	39	90	52
Partially translucent ($n = 21$)	7	68	10	61	48
Transition in color at the border ($p = 0.075$)					
Abrupt ($n = 40$)	32	47	23	60	43
Gradual ($n = 47$)	68	53	40	78	57
Color spectrum ($p = 0.007$)					
Monochromic ($n = 79$)	79	3	28	25	28
Oligochromic ($n = 8$)	21	97	75	72	72
Relation to microvasculature					
Predominant spacial relationship ($p = 0.014$)[b]					
Lesion fully obscuring vessels ($n = 37$)	68	47	46	69	56
Lesion partially obscuring vessels ($n = 19$)	16	61	21	52	43
Vessels crossing in front of lesion ($n = 7$)	16	92	57	63	62
Gross neovascularization ($p = 0.007$)[b]					
Absent ($n = 70$)	19	98	83	69	70
Present ($n = 6$)	81	2	31	17	30

PPV Positive predictive value; *NPV* negative predictive value

[a]Variables (test-positive) were compared with other variable(s) (test-negative) within the same group

[b]Excluded cases with no visible surrounding microvasculature

of only 74%, meaning at least one in four metastases would be missed.

Digital Image Processing

Measurement of Lesion Size In the short axis, the median measured size of benign lesions was 1.4 mm (range 0.3–7.5 mm) compared with 2 mm (range 0.6–5.1 mm) for metastases ($p = 0.015$). In the long axis, the median measured size of benign lesions was 2.3 mm (range 0.4–16.6 mm) compared with 3.4 mm (range, 0.8–8.8 mm) for metastases ($p = 0.011$). The median estimated surface area was 2.6 mm^2 (range 0.1–56.3 mm^2) for benign lesions and 5.9 mm^2 (range

TABLE 3 Risk score based on results from study cohort

Risk score[a]	Probability of harboring metastasis (%)	Accuracy (%)	No. of lesions needed for biopsy[b]
0	2	–	186
1	6	32	53
2	18	51	16
3	44	74	5
4	74	78	2

[a]Sum of points based on degree of nodularity (0 [flat], 1 [slightly nodular], 2 [markedly nodular]), color transition at the border 0 [abrupt], 1 [gradual]), and degree of transparency (0 [partially translucent], 1 [opaque])

[b]Calculated to have a 95% chance of at least one positive finding

0.4–31.7 mm^2) for metastatic lesions ($p = 0.013$). Reanalysis of the log transformed measurements of lesion size with GEE models to account for multiple lesions per person resulted in slightly higher p values (0.059 for the short axis, 0.105 for the long axis, and 0.072 for the area comparisons) of benign versus metastatic lesions.

Image Segmentation Color Analysis To determine any objectively measurable differences in color appearance of benign and malignant lesions, digital image segmentation analyses of all images were performed and provided the following results. The mean pixel intensity values of the ROI in the gray scale (as a surrogate for the combined color image) in each RGB color channel and each hue, saturation, value (HSV) channel did not differ statistically between benign lesions and peritoneal metastases (Table 4), indicating no difference in color appearance.

To correct for potential variability in the degree of illumination among different images, the mean values of the pixel intensity of the lesion's ROI were divided by the mean values of the pixel intensity of the lesion's ROI on the gray-scale image. This analysis again demonstrated no differences in results between benign lesions and peritoneal metastases (Table 4). Furthermore, the ROI of the lesions were compared with the immediate background surrounding the lesion by calculation of the signal-to-background ratio. These comparisons also showed no differences in measures of intensity between benign lesions and peritoneal metastases except for the hue (Table 4). The signal-to-background ratio of the hue was only 2.3 ± 3.2 for benign lesions versus 1.2 ± 1.6 for metastases ($p = 0.022$).

Measurement of Gradient Magnitude The median measured gradient magnitude was 113 ΔI(a.u.)/mm (range 6–806 ΔI(a.u.)/mm) for benign lesions versus 76 ΔI(a.u.)/mm (range 4–198 ΔI(a.u.)/mm) for metastases ($p = 0.019$, $p = 0.037$ from GEE of log transformed values). Yet the AUC was only 0.66 (95% CI 0.54–0.78), suggesting that this variable is a poor classifier. Furthermore, using a cutoff that provided the greatest Youden index (gradient magnitude, ≤ 140 ΔI(a.u.)/mm), this test would have missed at least one in seven metastases (i.e., 15%), and about half of benign lesions (i.e., 51%) would have been classified incorrectly as metastases, with an accuracy of only 61%.

Machine Learning

The findings showed no evidence that the applied DNN model was able to distinguish benign from metastatic lesions. The AUC was only 0.47 (95% CI 0.38–0.57), suggesting that the signal in these two classes of patches was too weak for the DNN model to detect, possibly due to the small sample size.

TABLE 4 Pixel intensity values in the ROI (grand mean ± standard deviation) separated by color channels from gray-scale, RGB, and HSV images

Image segments	Pixel intensity in lesion			Corrected pixel intensity in lesion[a]			Signal-to-background ratio[b]		
	Benign	Metastases	p Value[c]	Benign	Metastases	p Value[c]	Benign	Metastases	p Value[c]
Gray scale	177 ± 40	182 ± 26	0.506	–	–	–	1.3 ± 0.3	1.3 ± 0.3	0.896
Red channel	198 ± 38	203 ± 28	0.530	1.1 ± 0.1	1.1 ± 0.1	0.499	1.2 ± 0.2	1.2 ± 0.1	0.905
Green channel	168 ± 44	175 ± 28	0.467	0.9 ± 0.1	1.0 ± 0.1	0.264	1.4 ± 0.5	1.4 ± 0.3	0.913
Blue channel	164 ± 43	169 ± 29	0.598	0.9 ± 0.1	0.9 ± 0.1	0.975	1.6 ± 0.5	1.7 ± 1.2	0.576
Hue	115 ± 91	83 ± 77	0.106	0.7 ± 0.7	0.5 ± 0.4	0.066	2.3 ± 3.2	1.2 ± 1.6	0.022
Saturation	56 ± 33	51 ± 28	0.481	0.4 ± 0.3	0.3 ± 0.2	0.254	0.5 ± 0.1	0.6 ± 0.4	0.688
Value/brightness	197 ± 38	202 ± 28	0.518	1.1 ± 0.1	1.1 ± 0.1	0.522	1.2 ± 0.2	1.2 ± 0.1	0.464

ROI Region of interest; *RGB* red, green blue; *HSV* hue, saturation, value

[a]Calculated by lesion intensity divided by intensity of the entire image

[b]Calculated by lesion intensity divided by intensity in the immediate background

[c]p Values from two-sample student t tests (p values from generalized estimated equations [GEE] models were similar; data not shown)

DISCUSSION

In assessing the ability of an imaging test to detect peritoneal metastases in cancer patients, the key measure of quality is determined by the false-negative rate, typically measured by its sensitivity. Therefore, the applied test must have a false-negative rate as close as possible to 0% (i.e., sensitivity as close as possible to 100%) to avoid a clinically relevant rate of missed peritoneal metastases. The false-positive rate is clinically less important due to the typical low morbidity of peritoneal biopsies, but ideally, it should also be kept low. Strictly speaking, how many metastases are being missed is clinically more important than how many benign lesions are being biopsied mistakenly.

The results from this study suggest both that the false-negative rate achieved with routine staging laparoscopy is unacceptably great and that optical features able to help identify peritoneal metastases better are statistically relevant but clinically not sufficiently reliable. Specifically, even surgical experts demonstrated a clinically unacceptable rate of missed peritoneal metastases.

Given that the accuracy of optical predictors (e.g., degree of nodularity, transition in color at the border, and predominant degree of transparency) is inadequate, human performance is expected to remain limited, and decisions based on "gut feeling" are expected to continue providing substandard false-negative rates. Similarly, the computer performance in the form of the image segmentation analysis we used showed no difference in color appearance except for the hue, which also was clinically not reliable.

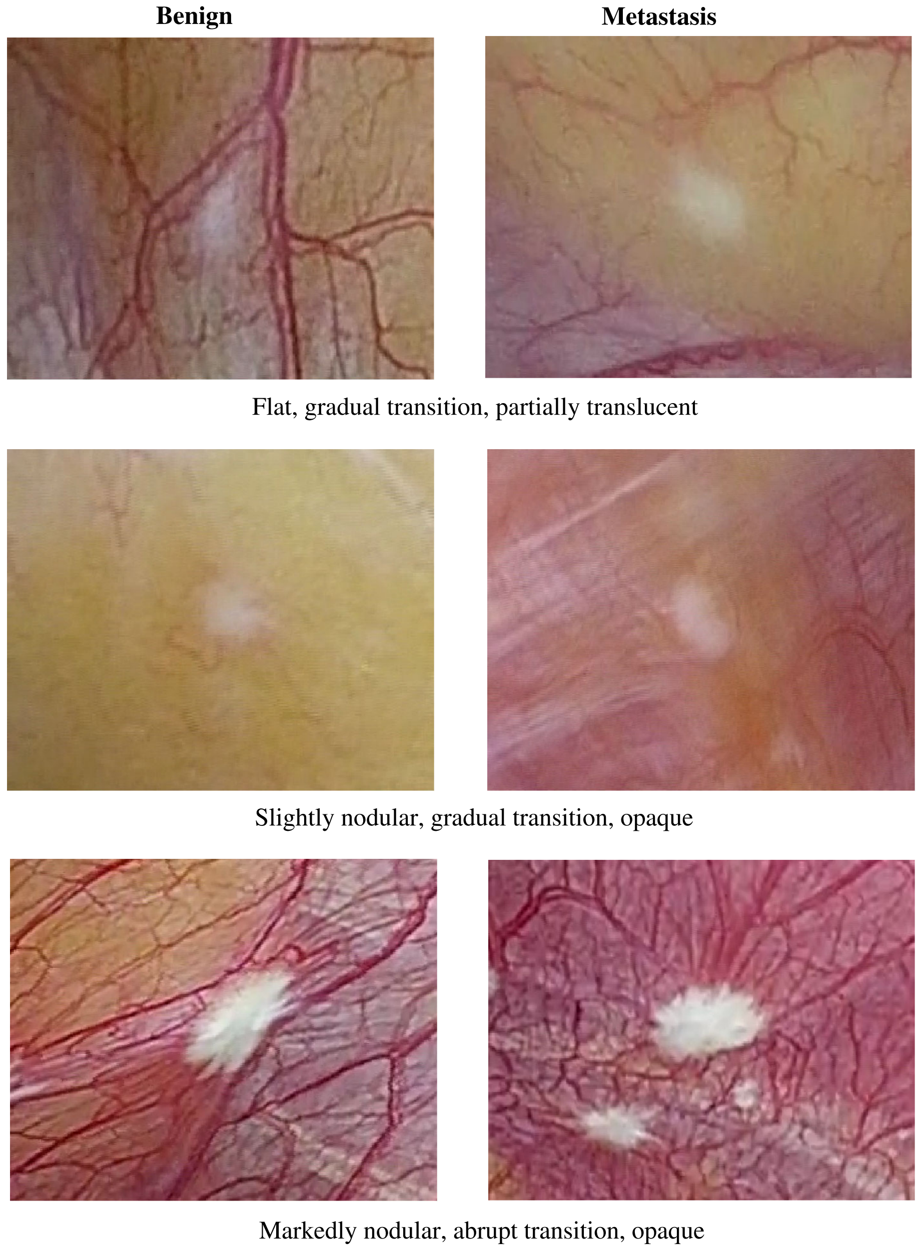
Of interest, computer image analysis, similar to the human performance, demonstrated that the transition in color at the border in the form of the gradient magnitude is seemingly a predictor but is only a poor classifier. Whereas machine learning comes with hopeful expectation for such a task, standard algorithms seem to provide a poor result at this point (although potentially related to the sample size given that machine learning usually requires very large samples to provide reliable results). Therefore, the only seemingly accepted conclusion from the results is that biopsy should be performed for any visible peritoneal lesion during cancer staging. This conclusion, however, comes with a pragmatic burden of significant added operative time.

A possible explanation for these findings is that laparoscopic color imaging, similar to direct visualization with the human eye, captures primarily information about the color and shape of the objects evaluated. Various subjective and objective measures were unable to determine features of color and shape that can reliably differentiate benign from malignant peritoneal lesions. This inability is likely because the pathophysiologic mechanisms involved in forming benign and malignant lesions can overlap (e.g., collagen scar formation from inflammation vs desmoplastic reaction from cancer). Because most cancer metastases are not a pure accumulation of cancer cells, similarities and therefore overlap in appearances between benign and malignant lesions are not surprising (Fig. 3). Standard imaging is likely not good enough for visualization of these differences, and some form of additional image-enhanced techniques/"optical biopsy" such as vector-enhanced imaging, fluorescence imaging, spectral imaging, confocal laser endomicroscopy, or optical coherence tomography would be needed if possible.[4]

The strength of our study was that it evaluated our hypothesis using multiple and different approaches, all supporting the same conclusion. Regarding limitations, there was an inherent selection bias by inclusion of only lesions deemed worthy of a biopsy at the time of the operation. Within the peritoneum of the study patients, additional lesions were noted during laparoscopy that were not biopsied due to a perceived benign appearance. Because the underlying pathology of these benign-appearing lesions is unknown, they could not be included. Such factors might potentially have influenced the findings but did not override them. Independently, this study clearly demonstrated that our average perception as well as any potentially available tool for assessing optical appearance is inadequate to classify peritoneal lesions reliably. Therefore, the selection bias should not have had a major impact on the conclusion of this study but could have affected the extent of the measured rates of false-negative clinical evaluations. The possibility of a false pathology result was believed to be very low and was not expected to have a substantial impact on the results. Another potential limitation involved visual inaccuracies of the involved human examiners. The visual inaccuracies of each examiner, including deficiencies in color vision/discrimination, were not evaluated in this study. Finally, our color analysis primarily applies to routine laparoscopy with its RGB photosensor and therefore is technically different yet practically similar to laparotomy with its light detection by human photoreceptors.

Therefore, although an independent statistical association was established between three variables and the presence of metastases, and although surgeons can identify metastases better than random selection, nevertheless, clinical concern maintains that using these three variables as part of a risk score model or using "gut feeling" does not reliably exclude the presence of peritoneal metastases at a clinically acceptable rate. The hope is that these results will help to identify additional metastases that otherwise would have been missed.

FIG. 3 Examples of selected peritoneal lesions from various patients demonstrating paired benign and malignant lesions with subjectively similar appearance



Although arguably everyone can agree that misinterpreting the extent of a malignancy cannot be a good thing, the potential implications of identifying otherwise missed peritoneal metastases have not been established at this point. Based on historic experiences, the current practice regarding operative resection of gastrointestinal malignancies is that if peritoneal (distant) metastases are grossly detectable (either radiographically or during an operative exploration), an operative resection usually is not thought to be beneficial except for significant palliation of symptoms. Although no definitive proof of concept exists, currently there is reasonable concern that the vast majority of patients with gastrointestinal malignancies experience the development of metastases at a very early stage of disease (typically before the diagnosis of the disease) and therefore harbor at the time of initial treatment, if not gross, at least micro-metastases (not all of which might present a life-limiting situation).[5,6] This assumption suggests that most R0 resections are indeed debulking operations. Therefore, it could be concluded that the role of operative resection/debulking is not so much dependent on the presence of metastases as on the extent of metastases together with the tumor biology. Historically, the cutoff for the extent had been measured by the presence of grossly detectable metastases. On the average, this cutoff seems reasonable. Therefore, it is imperative that all gross metastases be identified before an operative resection is performed for gastrointestinal malignancies, stressing the importance of the data from this study. For the future, however, this cutoff might encounter challenges once the aggressiveness of metastases can be measured better and once micro-metastasis can be detected adequately because the impact of operative debulking is thought to be best in the presence of metastases with limited extent, less aggressiveness, and good response to chemotherapy or biologic therapy. Therefore, adequate detection of metastases/cancer extent will always remain a crucial part of cancer care.

CONCLUSIONS

The implications from the findings of this study for current practice are that all visible peritoneal nodules/lesions need to be biopsied in the setting of a staging laparoscopy. If this nonselective approach is deemed not practical in a particular situation, at a minimum, the current "threshold" for biopsy of a peritoneal lesion needs to be lowered considerably. The problem surgeons still are expected to face with a selective approach, however, is that there are no reliable markers that allow adequate differentiation between benign and metastatic peritoneal lesions and therefore no definitive guidance on how to select and how to lower the "threshold." The risk score established in this study might be a starting point as a potential guide.

ACKNOWLEDGMENT The authors express gratitude to Brenda A. Joseph, CTR, and Jessica H. Miller, CTR (Cancer Registry, Lahey Health Cancer Institute) for their devoted assistance in identifying the study patients. They also thank all the gastrointestinal oncologic surgeons who volunteered their time by participating in the survey.

REFERENCES

1. Schnelldorfer T, Jenkins RL, Birkett DH, et al. Laparoscopic narrow band imaging for detection of occult cancer metastases: a randomized feasibility trial. *Surg Endosc.* 2016;30:1656–61.
2. https://keras.io. Accessed 20 Jan 2018.
3. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
4. Schnelldorfer T. Image-enhanced laparoscopy: a promising technology for detection of peritoneal micrometastases. *Surgery.* 2012;151:345–50.
5. Zhang ZY, Ge HY. Micrometastasis in gastric cancer. *Cancer Lett.* 2013;336:34–45.
6. Rhim AD, Mirek ET, Aiello NM, et al. EMT and dissemination precede pancreatic tumor formation. *Cell.* 2012;148:349–61.