Agreement between endoscopic ultrasound-guided fine-needle aspiration and endobiliary brush cytology in suspected pancreaticobiliary malignancies

Authors
Matthew J. Sullivan1, Hope Kincaid2, Shashin Shah1, Hiral N. Shah1

Institutions
1 Department of Medicine, Division of Gastroenterology and Hepatology, Lehigh Valley Health Network, Allentown, Pennsylvania, United States
2 Network Office of Research and Innovation, Lehigh Valley Health Network, Allentown, Pennsylvania, United States

submitted 6.5.2017
accepted after revision 24.7.2017

Bibliography
© Georg Thieme Verlag KG Stuttgart · New York
ISSN 2364-3722

Corresponding author
Matthew J. Sullivan, DO, Department of Medicine, Division of Gastroenterology and Hepatology, Lehigh Valley Health Network, 707 Hamilton Street, One City Center, 7th Floor, Allentown, PA 18101, United States
Fax: +1-484-862-3176
matthew.sullivan@lvhn.org

ABSTRACT

Background and study aims For suspected pancreaticobiliary malignancies, endobiliary brush cytology during endoscopic retrograde cholangiopancreatography (ERCP) remains the diagnostic test of choice despite historically poor and variable sensitivity. This has led to increased use of endoscopic ultrasound-guided fine-needle aspiration (EUS-FNA) as an initial test. Recently, our institution began using a cytology brush that was designed specifically to collect a more substantial and higher-quality sample. The aim of this study was to investigate whether this brush design would provide more adequate samples and have high agreement with EUS-FNA in patients who underwent both procedures.

Patients and methods A retrospective chart review was conducted of all patients who underwent both EUS-FNA and endobiliary brush cytology for suspicion of pancreaticobiliary malignancy from January 2013 to May 2015. A total of 41 patients met the inclusion criteria. Initially, sample quality was evaluated. Final cytology results were then assessed for agreement with EUS-FNA using Cohen’s kappa. The effect of considering atypical cytology as negative was also uniquely evaluated by running separate analyses.

Results Brush cytology provided an adequate sample in 95.1 % of cases. Cohen’s Kappa demonstrated moderate agreement between brush cytology and EUS-FNA: $\kappa = 0.42$ ($P=0.001$). When atypical results were excluded, agreement increased: $\kappa = 0.60$ ($P=0.02$), but remained moderate. If atypical results were considered “positive,” the two procedures demonstrated equal cancer detection rates of 80.8 %.

Conclusions The studied brush provided more adequate samples compared with historical rates for brush cytology and had moderate agreement with EUS-FNA. If this brush truly increases sample adequacy, it could potentially provide results comparable to EUS-FNA at lower cost.

Introduction
In the setting of a newly discovered biliary stricture or pancreatic mass, differentiating between a benign and malignant process is vital. Pancreatic adenocarcinoma and cholangiocarcinoma often present at an advanced stage owing to their nonspecific symptoms. This late presentation can render patients ineligible for treatment with curative intent [1, 2]. Therefore, obtaining a quick and accurate pathological diagnosis allows for optimal outcomes.

Multiple endoscopic diagnostic techniques are available for sampling a lesion of the pancreas or biliary system. The two most widely utilized are endobiliary brush cytology during endoscopic retrograde cholangiopancreatography (ERCP) and endoscopic ultrasound (EUS)-guided fine-needle aspiration (FNA) [3]. Endobiliary brush cytology was first described in 1975 [4], and the safety, ease, and ability to offer therapeutic
interventions at the time of biopsy have made it the initial diagnostic test of choice [5] and the standard of care in most institutions [6]. However, recent studies have concluded that EUS-FNA is superior to brush cytology [7, 8], with reported sensitivities of 60% to 93% [3]. However, not all endoscopists are formally trained in EUS, and therefore this technique might not be available at all institutions.

Specificity of 100% has been demonstrated with endobiliary brush cytology [9], yet reported sensitivities vary greatly from 4% to 65.8% [6, 10, 11]. The low sensitivity is most often attributed to a relatively high volume of false-negative diagnoses [12], related to sampling error, low cellular yield, and misinterpretation of the specimen owing to the often subtle differences between malignant and nonmalignant cells [13]. Location and type of tumor can also affect sensitivity, as pancreatic tumors can extrinsically compress the bile duct or even cause outlet obstruction, thereby limiting the number of cells obtainable by the brush [11, 14].

The traditional endobiliary brush designs used in earlier studies have a short segment (approximately 1.5 to 2 cm) of soft bristles oriented 45 to 90 degrees from the drive wire. Over time, innovative endobiliary brushes have been designed in attempts to improve sensitivity. One brush, which was designed with a tapered dilator and a 1 cm pad of Velcro with semi-rigid, mushroom-ended bristles, was shown to enhance tissue yield in a very small study [15], and another brush with a longer segment of more rigid bristles was found to significantly increase cellular yield, but not cancer detection rates [11].

A more recent product, the Infinity ERCP sampling device (US Endoscopy, Mentor, Ohio, USA), has been designed specifically with the goal of collecting a more substantial and high-quality sample than previous brush designs. The brush is available in two diameters, 2.5 mm and 3 mm, and is engineered with two types of bristles; the distal and proximal portions of the brush contain stiffer bristles designed to create a defect in the tissue, with the center portion comprising softer bristles to capture the abraded material. There are also spaces between the two types of bristles for collecting cells [16]. A picture of the brush can be seen in Fig. 1.

A 2014 retrospective review compared the Infinity ERCP sampling device against a traditional brush, and showed an increase in the average number of cell clusters and a cancer detection rate of 78% with the Infinity ERCP sampling device when following a standardized protocol. The authors concluded that the Infinity ERCP sampling device increases diagnostic accuracy, likely due to improved cellular yield [17]. However, a 2015 randomized controlled trial also compared the Infinity ERCP sampling device against a traditional brush in the setting of a suspected malignant biliary stricture, and found a higher cancer detection rate and sample quality with the traditional brush [18].

In 2013 our institution began the move from a traditional brush design to the Infinity ERCP sampling device. We hypothesized that use of the Infinity ERCP sampling device in suspected pancreatobiliary malignancies would provide more samples of adequate quality, thereby improving our cancer detection rate compared with historical data reported in the literature, and that the brush would have a high level of agreement with results of EUS-FNA. To evaluate our hypothesis, we conducted this retrospective study. Our aims were to investigate sample quality and to determine the level of agreement between the Infinity ERCP sampling device and EUS-FNA in patients who underwent both procedures for a suspected malignant pancreatobiliary lesion.

**Patients and methods**

**Patient characteristics and data collection**

A retrospective chart review was conducted of over 300 patients who underwent ERCP with endobiliary brush cytology for any indication between 1 January 2013 and 1 May 2015 at Lehigh Valley Hospital – Cedar Crest in Allentown, Pennsylvania, USA. The study was approved by the Institutional Review Board of Lehigh Valley Health Network. Requirement of informed consent was waived at the time of approval due to the retrospective design.

For inclusion in the study, patients must have undergone endobiliary brush cytology for a newly suspected pancreatobiliary malignancy; brush cytology must have been performed with the Infinity ERCP sampling device, and patients must have also previously or subsequently undergone EUS-FNA for the same indication. FNA was performed using the EchoTip FNA needle, EchoTip ProCore fine needle biopsy system (Cook Medical, Bloomington, Indiana, USA), or the SharkCore fine needle biopsy system (Medtronic, Minneapolis, Minnesota, USA) in either 22 or 25 gauge.

Suspicion of malignancy was based on abnormal hepatobiliary-associated laboratory tests, including aspartate transaminase, alanine transaminase, alkaline phosphatase, and total bilirubin, a mass seen on previous imaging, and/or clinical presentation and course. Patients with a previously known diagnosis of pancreatobiliary malignancy were excluded.

All procedures were performed by one of two gastroenterologists who are formally fellowship-trained in advanced endoscopy. Strictures were brushed with the Infinity ERCP sampling device using a minimum of seven up-and-down motions before the brush and catheter were removed together. The obtained cells were collected in cytology solution, and the entire brush head was clipped and included in the sample. The brush catheter was then flushed into the same cytology solution and the entire sample was submitted for cytology. All samples were interpreted by a group of in-house multidisciplinary pathologists with gastroenterology experience.
First, sample quality was deemed adequate, less than optimal, or inadequate for evaluation by the interpreting pathologist. Our pathologists made this determination primarily based on the cellularity of the sample. Samples deemed indeterminate by our pathologists were reviewed in consultation with a national expert gastroenterology cytopathologist at an outside institution, and the interpretation by the outside pathologist was used as the final result. All adequate and less-than-optimal samples were then classified into one of four categories: diagnostic for malignancy, suspicious for malignancy, atypical cellularity, or negative for malignancy. There are no standard cytopathological features that pathologists use to classify biliary cytology samples; categorization of the samples in our study was based on level of cellularity and presence or absence of characteristic features of malignancy as determined by the interpreting cytopathologist. Examples of the four categories are shown in Fig. 2. In our study, suspicious for malignancy was considered positive based on the high positive predictive value for malignancy of this category [19].

The final diagnosis was determined by either the cytological diagnosis obtained during the study procedures plus surgical pathology, radiographic imaging revealing metastatic disease, or the diagnosis documented by an oncologist at a follow-up appointment or listed on the death certificate of those patients who were deceased. Follow-up data were also obtained by retrospective chart review and were available for a range of 26 to 52 months depending on the date of the initial procedure.

**Statistical analysis**

Descriptive statistics were used to summarize the characteristics of the study sample. The mean and standard deviation are reported for continuous variables that were found to be normally distributed, as assessed by the skewness statistics and visual inspection of histograms. The median, interquartile range, and range are reported for continuous variables that were not normally distributed.

The sensitivity and specificity were calculated for both procedures assuming different categorizations of atypical results. The 95% confidence interval (CI) is presented alongside each estimate. Cohen’s kappa was used to assess the level of agreement between diagnostic results of the two procedures. A value of <0.20 is considered poor agreement, 0.21–0.40 is fair agreement, 0.41–0.60 is moderate agreement, 0.61–0.80 is good agreement, and 0.81–1.00 is very good agreement.

All analyses were conducted using SAS version 9.3 (SAS Institute, Cary, North Carolina, USA). The 95% CIs and P values reported were calculated using the exact method given the small sample size. A P value of <0.05 was considered to be statistically significant.

**Results**

From 1 January 2013 to 1 May 2015, 41 patients underwent both brush cytology with the Infinity ERCP sampling device and EUS-FNA for a newly suspected pancreaticobiliary malignancy. Overall, 15 of the 41 patients had core tissue sampling during EUS performed with a fine-needle biopsy system. The majority of patients (78.1%) underwent both procedures during the same endoscopy session. Of the remaining patients, all but one underwent the second procedure within 14 days of the first procedure. In one patient, the procedures were 73 days apart. Patient characteristics are listed in Table 1. Malignancy was ultimately diagnosed in 26 patients (63.4%), with pancreatic adenocarcinoma being the most prevalent (76.9%). The remaining 15 patients had no malignancy identified on either procedure: 2 had strictures due to primary sclerosing cholangitis, 2 due to chronic pancreatitis, 1 was felt to be due to previous radiation, 8 were determined to be benign reactive strictures, 1 patient was diagnosed with high grade dysplasia of the bile duct, and 1 patient had atypical cellularity with the inability to exclude malignancy and continues to be followed.

The Infinity ERCP sampling device provided an adequate sample in 39 of the 41 patients (95.1%), with one additional sample (2.4%) deemed less than optimal but still analyzable. For EUS-FNA, 33 of 41 samples were adequate (80.5%), four were less than optimal (9.8%), and another four were inadequate for evaluation (9.8%). All of the inadequate samples from EUS were obtained with a standard FNA needle. All patients had an adequate sample from at least one of the procedures.
Brush cytology resulted in 15 samples considered diagnostic for malignancy (36.6%) and EUS-FNA resulted in 18 (43.9%). In the 26 patients who were ultimately diagnosed with malignancy, both methods had been diagnostic for malignancy in 11 – 8 cases of pancreatic adenocarcinoma and 3 of cholangiocarcinoma. In two patients, both studies diagnosed atypical cellularity; one of these patients was ultimately diagnosed with pancreatic adenocarcinoma and the other without an identified malignancy. Both studies were negative for malignancy in eight patients, but one patient was eventually diagnosed with hepatic angiosarcoma by liver biopsy. One brush cytology sample was suspicious for malignancy; however, no FNA results were classified as suspicious for malignancy, and therefore the agreement between samples of this classification could not be assessed and that sample was excluded from the analysis.

For patients in whom both procedures returned adequate samples, Cohen’s kappa showed a moderate level of agreement between the two procedures: $\kappa = 0.42 \ (P = 0.001)$. When samples categorized as atypical cellularity and suspicious for malignancy were excluded, agreement increased but remained at a moderate level ($\kappa = 0.60; \ P = 0.02$).

In 11 patients, one procedure was diagnostic for malignancy but the other provided a different result. Of those, five had atypical cellularity from the other procedure and all were ultimately diagnosed with malignancy – four pancreatic adenocarcinomas and one of uncertain histogenesis. Three patients had a brush cytology result that was negative for malignancy while FNA was diagnostic for malignancy, and all three were diagnosed with pancreatic adenocarcinoma. Similarly, two patients had an FNA result that was negative for malignancy while brush cytology was diagnostic for malignancy – one pancreatic adenocarcinoma and one cholangiocarcinoma. There was also one case in which brush cytology results were diagnostic for malignancy and the FNA sample was inadequate for analysis. This patient was found to have high grade dysplasia of the bile duct, but no malignancy. ▶ Fig. 3 and ▶ Fig. 4 demonstrate the results of comparison of the samples obtained from brush cytology and EUS-FNA in flow charts.

### Table 1 Background characteristics (n = 41).

<table>
<thead>
<tr>
<th>Age, mean (SD), years</th>
<th>70.4 (10.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>20 (48.8)</td>
</tr>
<tr>
<td>Female</td>
<td>21 (51.2)</td>
</tr>
<tr>
<td>Mass on prior imaging$^1$, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>15 (36.6)</td>
</tr>
<tr>
<td>No</td>
<td>26 (63.4)</td>
</tr>
<tr>
<td>Lab results$^2$, n (%)</td>
<td></td>
</tr>
<tr>
<td>Abnormal</td>
<td>33 (80.5)</td>
</tr>
<tr>
<td>Normal</td>
<td>1 (2.4)</td>
</tr>
<tr>
<td>Not available</td>
<td>7 (17.1)</td>
</tr>
</tbody>
</table>

$^1$ Prior imaging included computed tomography, magnetic resonance imaging, or abdominal ultrasound.

$^2$ Laboratory tests included aspartate transaminase, alanine transaminase, alkaline phosphatase, and total bilirubin.

---

▶ Fig. 3 Comparison of results from brush cytology and endoscopic ultrasound-guided fine-needle aspiration when the two procedures were in agreement.
Fig. 4. Comparison of results from brush cytology and endoscopic ultrasound-guided fine-needle aspiration (FNA) when the two procedures did not agree.
We also explored the effect of considering atypical results as “negative” by running separate analyses—one where atypical results were considered “positive” and a traditional analysis in which they were considered “negative.” Brush cytology resulted in 10 samples with atypical cellularity. If considered “positive” 65.0% of patients would have positive cytology results, and if considered “negative” 40.0% would have positive cytology results. EUS-FNA resulted in five samples with atypical cellularity. If considered “positive” 62.2% of patients would have positive cytology results, and if considered “negative” 48.7% would be positive.

After running the analyses with dual consideration of atypical results and including only those samples that were deemed adequate by both procedures, agreement between the procedures remained moderate by Cohen’s kappa in both scenarios. When atypical was considered “positive” $\kappa = 0.41 \ (P = 0.04)$, and when considered “negative” $\kappa = 0.49 \ (P = 0.01)$.

Sensitivity, specificity, positive predictive value, and negative predictive value were also reported with dual consideration of atypical cellularity. Cancer detection rates were also calculated. These represent the sensitivity of the procedure with inadequate samples considered negative in order to emulate real-world practice; therefore, inadequate samples are included in this calculation. Results are reported in Table 2.

**Table 2** Diagnostic characteristics of both procedures under different categorizations of atypical cellularity samples and excluding inadequate samples.

<table>
<thead>
<tr>
<th></th>
<th>Brush cytology (Atypical negative)</th>
<th>Brush cytology (Atypical positive)</th>
<th>EUS-FNA (Atypical negative)</th>
<th>EUS-FNA (Atypical positive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity, % (95 %CI)</td>
<td>60.0 (38.7 – 78.9)</td>
<td>84.0 (63.9 – 95.5)</td>
<td>69.2 (48.2 – 85.7)</td>
<td>80.8 (60.7 – 93.5)</td>
</tr>
<tr>
<td>Specificity, % (95 %CI)</td>
<td>93.3 (68.1 – 99.8)</td>
<td>66.7 (38.4 – 88.2)</td>
<td>100 (71.5 – 100)</td>
<td>81.8 (48.2 – 97.7)</td>
</tr>
<tr>
<td>PPV, % (95 %CI)</td>
<td>93.8 (69.8 – 99.8)</td>
<td>80.8 (60.7 – 93.5)</td>
<td>100 (81.5 – 100)</td>
<td>91.3 (72.0 – 98.9)</td>
</tr>
<tr>
<td>NPV, % (95 %CI)</td>
<td>58.3 (36.6 – 77.9)</td>
<td>71.4 (41.9 – 91.6)</td>
<td>57.9 (33.5 – 79.8)</td>
<td>64.3 (35.1 – 87.2)</td>
</tr>
<tr>
<td>Cancer detection rate, %</td>
<td>57.7</td>
<td>80.8</td>
<td>69.2</td>
<td>80.8</td>
</tr>
</tbody>
</table>

PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval; EUS-FNA, endoscopic ultrasound-guided fine-needle aspiration.

1 There was one inadequate brush cytology sample and four inadequate EUS-FNA samples. Therefore, calculations were performed using 40 samples for brush cytology and 37 for EUS-FNA.

Categorization of brush cytology samples can also influence the sensitivity of brush cytology, as there are no well-established objective cytomorphological criteria for diagnosing biliary brushings [19]. Premalignant lesions of the biliary tract are not as well defined compared with other malignancies, and therefore distinguishing low grade dysplasia from reactive changes may not always be possible because of inter- and intraobserver variability [12]. A more experienced or “aggressive” cytopathologist may diagnose malignancy with fewer abnormal cells or characteristic changes, while more conservative cytopathologists may require more definitive changes and higher cellular yield [11]. By providing a higher-quality sample, the Infinity ERCP sampling device may help limit interpretation error and improve sensitivity.

Traditionally, studies of brush cytology have considered atypical cellularity to be “negative” [6,12,17,20]. Our study attempted to uniquely evaluate this potential confounder by running separate analyses with dual consideration of atypical cellularity samples. While we recognize that consideration of atypical cellularity as “positive” goes against what is commonly accepted, by doing so we demonstrated that categorization of atypical results, especially in the setting of relatively small sample sizes, could explain the wide-ranging sensitivities of brush cytology reported in previous studies.

We also demonstrated only a moderate level of agreement between the two procedures, even when atypical cellularity was excluded. As EUS-FNA continues to gain traction as a diagnostic procedure in this setting, and without a standardized protocol, choosing a procedure that will result in an accurate diagnosis while also limiting morbidity and cost to the patient and health care system should also be taken into consideration. Not all endoscopists are trained in EUS and therefore this technology may not be as widely available as brush cytology. Also, FNA cannot always be performed directly on a mass or biliary stricture, resulting in biopsy of adjacent structures such as lymph nodes. This can potentially provide staging data, but benign results of biopsies from these areas may represent an unnecessary procedure resulting in increased risk of morbidty to the patient. It also results in increased cost, as one Infinity ERCP sampling device costs our institution $95, while one SharkCore...
fine-needle biopsy system costs $240. Interestingly, the Infinity ERCP sampling device also costs our institution less than the traditional model that we previously utilized, which is priced at $103. Also, in analyzing cost, a majority of our patients underwent both ERCP and EUS during the same session. When factoring in resources of the endoscopy, laboratory, anesthesia services, and potential repeat procedures, performing both studies may potentially be more cost-effective if there is a lack of certainty following EUS-FNA.

There have also been reports of tumor seeding, or spread of malignancy along a biopsy tract, during EUS-FNA [21–24], which has not been demonstrated to occur during brush cytology [10]. Although a rare occurrence, this is something that must be taken into consideration for possible cholangiocarcinoma, where prior FNA can be a contraindication to transplantation due to the possibility of peritoneal seeding [25].

Additionally, we would like to highlight three cases—two pancreatic adenocarcinomas and one cholangiocarcinoma—where the Infinity ERCP sampling device diagnosed malignancy and EUS-FNA was either negative for malignancy or provided atypical cellularity. If brush cytology was not performed, these malignancies may have been missed or additional diagnostic procedures at increased cost and risk of morbidity would have been required.

Limitations of our study include the retrospective nature and small sample size. Our inclusion criteria were less strict than similar studies; however, we excluded patients with known malignancy, which likely decreased our pre-test probability of malignancy. Our institution was also undergoing a transition to a new fine-needle biopsy system during the time of this study, so not all FNA cytology samples were obtained from the same device. A strength of our study is that all patients underwent both procedures, reducing the impact of potential confounders due to patient characteristics.

In conclusion, we found that the Infinity ERCP sampling device resulted in increased sample adequacy compared with historical rates of brush cytology; this correlates with previous studies of this brush. For patients in whom both procedures returned adequate samples, Cohen’s kappa revealed a moderate level of agreement between the results of brush cytology and EUS-FNA: κ = 0.42 (P = 0.001). If the Infinity ERCP sampling device truly increases sample adequacy it could potentially provide results comparable to EUS-FNA at a lower cost. Larger studies are needed to further validate these findings. Future studies to determine a standardized protocol for evaluation of suspected pancreaticobiliary malignancy are also needed. In the interim, we recommend continuing to rely on pre-test probability with use of both EUS-FNA and brush cytology in evaluation of suspected pancreaticobiliary malignancies.

Competing interests

Dr. Shah is currently on the speaker’s bureau for Abbvie. He is also a former speaker for US Endoscopy and Olympus.

References


