Factors Impacting the Performance Characteristics of Bile Duct Brushings

A Clinico-Cytopathologic Analysis of 253 Patients

Ezgi Hacihasanoglu, MD; Bahar Memis, MD; Burcin Pehlivanoglu, MD; Vaidehi Avadhani, MD; Alexa A. Freedman, BA; Michael M. Goodman, MD, MPH; N. Volkan Adsay, MD; Michelle D. Reid, MD, MSc

• ***Context.***—**Literature on factors impacting bile duct brushings (BDBs) performance characteristics remain limited.**

Objective.—**To capture the current state of daily practice with BDB sign-out.**

Design.—**Two hundred fifty-three of 444 BDBs signed out by more than 7 cytopathologists, with histopathologic and/or clinical follow-up of at least 18 months, were examined.**

Results.—**One hundred thirty-five of 253 BDBs (53%) had histologically confirmed malignancies, 22 (9%) had cancer-related deaths, and 96 (38%) were benign. Cytologic diagnoses in the 444 BDBs were nondiagnostic (11 [2.5%]), negative (284 [64%]), atypical (62 [13.9%]), suspicious (34 [7.7%]), and malignant (53 [11.9%]). Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy of malignancy detection were 35%, 100%, 100%, 58%, and 66%, respectively. When atypical, suspicious, and malignant (ASM) categories were combined, sensitivity increased (58%), specificity and PPV dropped (97%), and accuracy increased (73%). Carcinoma type (bile-duct versus pancreatic-ductal) had no effect on accuracy ($P = .60$) or diagnostic class ($P = .84$), nor did time of performance (first 7.5 versus latter 7.5 years, $P = .13$). Interestingly, ThinPrep + cell block (n = 41) had higher sensitivity (61%) and lower specificity (80%) than ThinPrep only (versus 51% and 100%, respectively). Sensitivity and specificity were higher (47% and 100%) in nonstented than stented specimens (59% and 97%). Relative risk of malignancy for "suspicious" (2.30) and "atypical" (2.28) categories was lower but not very different from that of "malignant" category (2.41).**

Conclusions.—**Bile duct brushings had fairly low sensitivity but high specificity and PPV with no false positives. Sensitivity almost doubled and specificity dipped minimally when ASM categories were combined, highlighting the need for better classification criteria for atypical/suspicious cases. Higher specificity, PPV, NPV, and accuracy but lower sensitivity in stented BDBs suggest that they be called malignant only when evidence is overwhelmingly convincing.**

(*Arch Pathol Lab Med.* 2018;142:863–870; doi: 10.5858/arpa.2017-0150-OA)

The diagnosis of malignancy in the biliary tree remains a clinical and pathologic challenge, with brush cytology being the most frequently used initial diagnostic test. Cytologic diagnosis is especially challenging, as tumors are often well differentiated and thus cytologically bland. Conversely, these well-differentiated examples are closely mimicked morphologically by reactive changes in patients with primary sclerosing cholangitis, nonspecific inflammation, cholelithiasis, and stents.[1,2]

Most studies on bile duct brushings (BDBs) agree that the sensitivity of this modality is fairly low, although the numbers vary substantially from 6% to 64%,[3–16] and the factors implicated in the test's inaccuracies are also significantly different. These variations are presumably related to case definition (identification) as well as the setting in which the studies were performed and the different methods of specimen processing used.[3,17,18] Consequently, the relative impact of different parameters on BDB interpretation by cytopathologists remains unclear and is yet to be fully unraveled.

In this study, in order to capture the current state of daily practice, we reviewed our institution's experience with the cytomorphologic analysis of BDBs during the past 15 years to further elucidate the factors that impacted the performance characteristics of the test. Since 11 cytopathologists reviewed the cases during the 15-year period, the potential

Accepted for publication October 11, 2017.

Published as an Early Online Release March 27, 2018.

From the Department of Pathology, Emory University School of Medicine, Atlanta, Georgia (Drs Hacihasanoglu, Memis, Pehlivanoglu, Avadhani, and Reid); the Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia (Ms Freedman and Dr Goodman); and the Department of Pathology, Medical College of Wisconsin, Milwaukee (Dr Adsay).

The authors have no relevant financial interest in the products or companies described in this article.

This research was presented as an abstract presentation at the 105th Annual Meeting of the United States and Canadian Academy of Pathology (USCAP); March 2016; Seattle, Washington.

Corresponding author: Michelle D. Reid, MD, MSc, Department of Pathology and Laboratory Medicine, Emory University Hospital, 1364 Clifton Rd NE, Room G179B, Atlanta, GA 30322 (email: michelle.reid@emory.edu).

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2017-0150-OA by guest on 07 January 2022

impact from observer variability and bias could not be eliminated and so to address this, an additional blinded review and diagnosis of a subset of 60 BDBs (30 benign and 30 malignant) was performed by 7 select reviewers. These results are outlined in greater detail in a previously published article.[19]

MATERIALS AND METHODS

Study Setting

The study was approved by the Emory University Hospital (Atlanta, Georgia) Institutional Review Board. Bile duct brushings had been obtained by multiple different gastroenterology groups with different levels of expertise, and the diagnostic interpretation of the cytology material was provided by 11 different cytopathologists (all board-certified and >95% were signed out by 7 cytopathologists) at our institution's Pathology Department, which serves both as a primary care center for the local communities as well as a tertiary center facility with pathology residency and cytopathology fellowship programs.

Materials and Criteria for Study Inclusion

A search for all BDB samples between 2000 and 2015 yielded 444 patients for whom clinicopathologic data (age, sex, race), as well as information regarding presence or absence of stents, BDB preparation method, cytologic diagnoses and (where available) histologic diagnoses on concurrent or subsequent biopsies or resections, and clinical outcome data were collected. To avoid patient duplication, for patients who had more than 1 brushing during the 15-year period, the first chronologic brushing was used as their index sample and their follow-up brushings were not included in statistical analysis, which was limited to only index patient samples.

As a part of the official diagnostic report generated at Emory University Hospital's Cytopathology Section, each case had been formally classified specifically as "malignant," "suspicious," "atypical," and "negative for malignant cells." This allowed documentation of specific classification of each case, based entirely on the original cytopathology reports, without necessitating that the authors make inferences from the various terminology used by the original interpreters who established the specific categorical diagnoses. However, for the "atypical" group, there were different "tiers" of concern expressed in the comment section of the reports, which were divided into 3 categories as "atypical, favor reactive," "atypical, not otherwise specified (NOS)," and "atypical, cannot rule out carcinoma." For the purpose of statistical analysis all "atypical, favor reactive" diagnoses were combined with the "negative for malignant cells" category. "Atypical, NOS" and "atypical, cannot rule out carcinoma" were combined and analyzed as the "atypical" category. Cytologic features used to evaluate and classify a sample as malignant included the presence of hypercellularity, 3-dimensional clusters, nuclear pleomorphism, high nuclear to cytoplasmic ratio (>2:1), nuclear irregularity, prominent nucleoli, nuclear molding/hugging, single discohesive cells with high nuclear to cytoplasmic ratio, change in chromatin pattern, cytoplasmic mucin vacuoles, and a 2-cell population. The analysis of samples was purely morphologic, and ancillary studies (such as fluorescence in situ hybridization and immunohistochemistry) were not used to evaluate samples.

Final Outcome

A patient's final outcome was considered a benign or malignant outcome and this information was used as the gold standard for follow-up. Patients were considered as having a malignant outcome if malignancy was diagnosed on concurrent or subsequent biopsy, fine-needle aspiration, or resection of a primary or metastatic pancreatobiliary tract tumor. Patients were also determined to have a malignant outcome if despite not having histologic confirmation, they developed unequivocal clinical signs of malignancy and were given oncologic treatment or died of malignancy.

Patients were considered as having a benign outcome if they had follow-up pancreatobiliary resections that were benign or if they had an uneventful clinical follow-up period of 18 or more months. Unlike many previous studies, which used 6 months of follow-up as proof of benignity,[18] in this study, patients with fewer than 18 months of uneventful follow-up were not included in statistical analysis because, although cholangiocarcinoma has an average survival of 18 to 24 months and virtually all patients develop definitive signs of malignancy in that time frame, a not insignificant percentage of cases demonstrate their malignant nature within 1 to 1.5 years, and thus, following up patients for fewer than 18 months would miss numerous more indolent tumors.[20,21]

Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy of BDB in identifying malignancy were determined by comparing the cytologic diagnosis to the gold standard of the final follow-up clinical outcome. Samples classified as malignant, suspicious, or atypical (atypical, NOS; and atypical, cannot rule out carcinoma) were considered to have a "positive" cytologic diagnosis. These statistics were also calculated separately for each preparation method, presence or absence of a stent, and type of carcinoma. Additionally, relative risks (RRs) were calculated to determine the risk of a malignant final outcome associated with each of the cytologic diagnosis classifications. Since 11 cytopathologists reviewed the cases during a 15-year period, the potential impact from observer variability and bias could not be eliminated. To address this impact, an additional, blinded review and diagnosis was performed by 7 select reviewers on a select subset of 60 BDBs (30 benign and 30 malignant). These reviewers were blinded to clinical information, cytologic diagnosis, and follow-up outcome of the patients. Detailed results of this blinded review were previously published and presented in detail in a separate article.[19]

RESULTS

Of the 444 patients who had undergone BDBs between 2000 and 2015, there were 191 patients excluded from analysis because of incomplete or missing follow-up information. This left 253 patients (57%) who met criteria for benign or malignant outcome. Of these, there were 120 females and 133 males. The mean patient age was 61 years (range, 19–95 years). There were 101 African American, 128 white, and 4 Asian patients. Race was unknown in 20 patients. Of these 253 patients, 157 (62%) had malignant outcome (including 135 histologically confirmed cancers and 22 cancer-related deaths) and 96 (38%) had benign outcome (proven either by subsequent resection showing benign disease or by ≥18 months of uneventful follow-up). A stent was present at the time of BDB in 43 of 253 patients (17%), while 210 of 253 patients (83%) did not have a stent at the time of BDB.

Cytologic Diagnoses

Cytologic diagnostic categories used for all BDBs were as follows: inadequate, benign, atypical cells, suspicious for malignancy, and malignant. The atypical category was further divided into 3 groups: "atypical, favor reactive," "atypical, NOS," and "atypical, cannot rule out carcinoma." Cytologic diagnoses for the 253 patients were as follows: nondiagnostic for 6 (2%), negative for malignant cells for 124 (49%), atypical for 66 (26%), suspicious for malignancy for 22 (9%), and malignant for 35 (14%) (Table 1). There were 29 (11%) "atypical, favor reactive"; 23 (9%) "atypical, NOS"; and 14 (6%) "atypical, cannot rule out carcinoma." Some of these cases are illustrated in Figures 1, A and B; 2, A and B; 3, A and B; and 4, A and B.

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2017-0150-OA by guest on 07 January 2022

Table 1. Cytologic Diagnoses for Patients With Follow-up (n = 253)		
Diagnostic Category	**No.**	**%**
Nondiagnostic	6	2
Benign	124	49
Atypical, favor reactive	29	11
Atypical, not otherwise specified	23	9
Atypical, cannot rule out carcinoma	14	6
Suspicious for malignancy	22	9
Malignant	35	14

Patients With Malignant Outcome

Of the 253 patients, 135 (53%) had histologically confirmed malignancy, 69 (51%) of which were diagnosed on resection and 66 (49%) on biopsy; 22 additional patients had malignant outcome based on clinical follow-up with 19 cancer-related deaths, 2 with metastases, and 1 with inoperable, locally advanced disease.

The diagnostic procedures for patients with malignancy included pancreatoduodenectomy (n = 57), total/segmental hepatectomy (n = 7), cholecystectomy (n = 3), bile duct resection (n = 2), and biopsy (n = 66). The 135 histologically confirmed malignancies included pancreatic ductal adenocarcinoma (n = 87, 64.5%), bile duct carcinoma (n = 41, 30.4%), gallbladder carcinoma (n = 3, 2.2%), neuroendocrine tumor (n = 2, 1.5%), hepatocellular carcinoma (n = 1, 0.7%), and metastatic colonic adenocarcinoma (n = 1, 0.7%) (Tables 2 and 3).

Patients With Benign Outcome

A total of 96 patients (38%) had benign outcome. Of these, 25 (26%) had resections that were negative for tumor and 71 (74%) had at least 18 months of uneventful clinical follow-up.

Overall Bile Duct Brushing Cytology Performance

The performance characteristics of bile duct brush cytology were determined for all specimens including those

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2017-0150-OA by guest on 07 January 2022

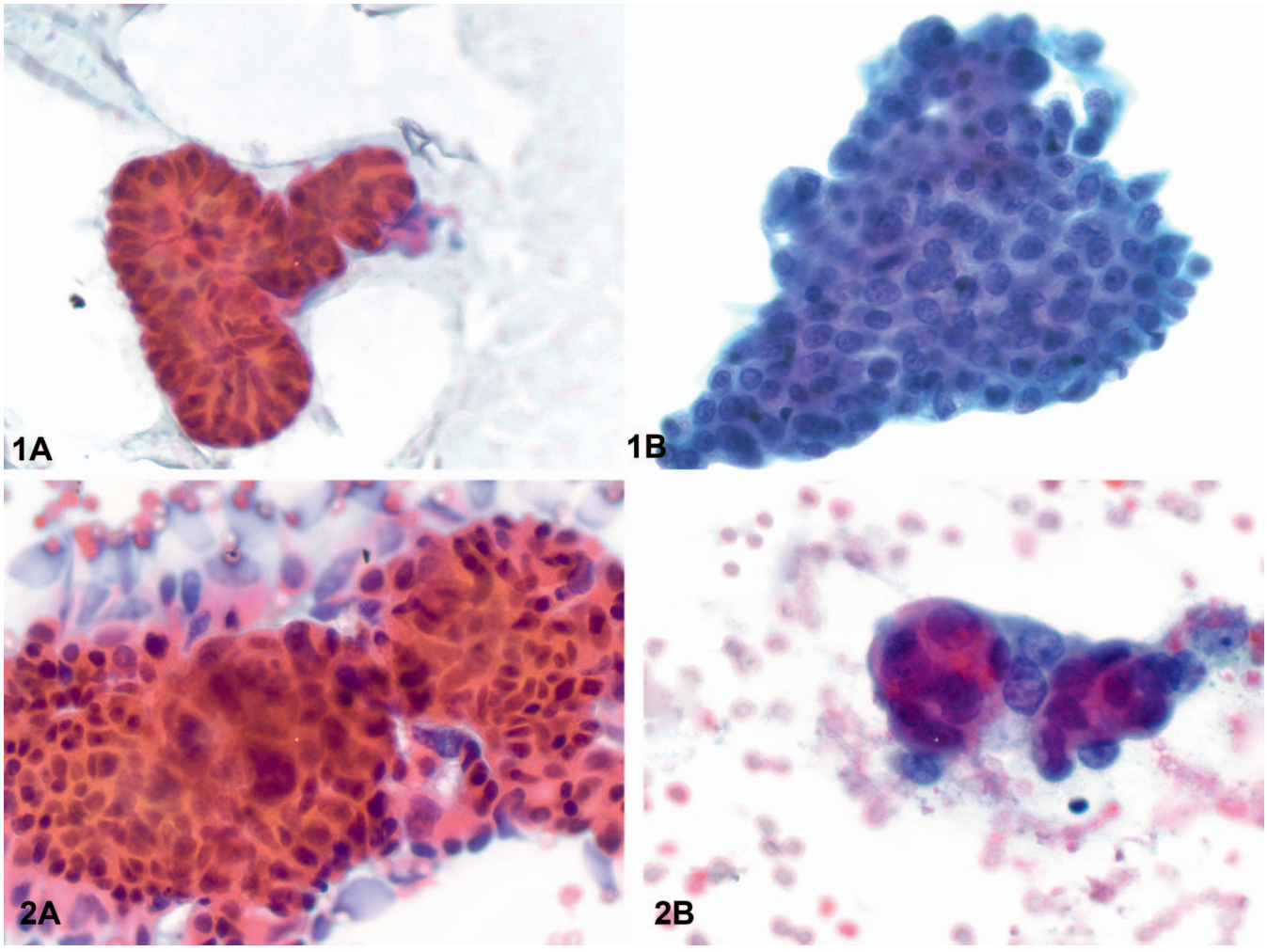


Figure 1. *A, Smear of a benign bile duct brushing showing classical features of reactive "detachment" atypia with folded comma-shaped sheet of benign epithelial cells with slightly increased nuclear to cytoplasmic ratio and uniformly distributed chromatin. B, ThinPrep of a benign bile duct brushing showing flat honeycomb sheet of epithelial cells with low nuclear to cytoplasmic ratio, indistinct nucleoli, and evenly distributed chromatin (Papanicolaou stain, original magnifications ×200 [A] and ×400 [B]).*

Figure 2. *A, Smear of a bile duct brushing that, despite air-drying artifact, highlights drunken honeycomb sheet of malignant cells with nuclear pleomorphism, high nuclear to cytoplasmic ratio, and coarse chromatin. B, High-power view of same case in (A) showing 3-dimensional cell cluster with nuclear pleomorphism and irregular nuclear borders (Papanicolaou stain, magnifications ×200 [A] and ×400 [B]).*

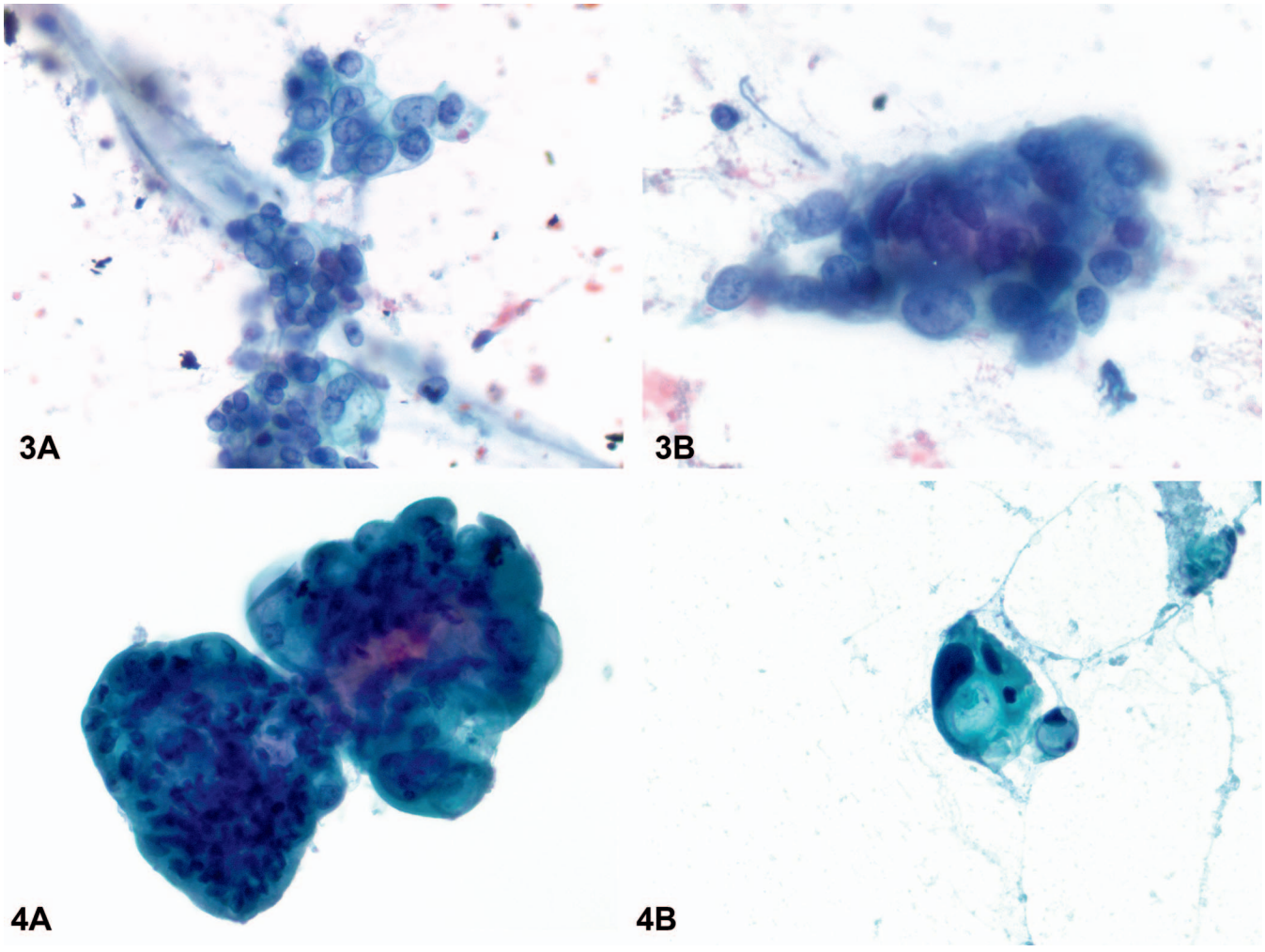


Figure 3. *Smear of a bile duct brushing showing malignant cells with pale cytoplasm, striking hypochromasia, cytoplasmic mucin vacuoles, and cellular discohesion (A) with rare pleomorphic 3-dimensional clusters (B) (Papanicolaou stain, magnification ×200 [A] and ×400 [B]).*

Figure 4. *A, ThinPrep of a bile duct brushing showing malignant cells with voluminous cytoplasm, overlapping nuclei with prominent nucleoli, and marked nuclear pleomorphism with focal 4-fold anisonucleosis. B, Same case as (A) showing a hyperchromatic malignant cell cluster with prominent cytoplasmic mucin vacuoles and clinging necrosis (Papanicolaou stain, magnification ×400 [A and B]).*

Table 2. Factors Affecting Test Performance in Patients With Known Follow-up (n = 253)

	%	Sensitivity, %	Specificity, %	PPV, %	NPV, %	Accuracy, %
Cytology diagnosis (n)						
Malignant (n = 35)	14	35	100	100	58	66
Suspicious (n = 22)	9	24	95	95	58	63
Atypical[a] (n = 37)	15	35	98	95	58	65
Malignant + suspicious + atypical (n = 94)	37	58	97	97	58	73
Preparation method (n)						
Smears (n = 29)	18	67	100	100	74	83
Smears + cytospin (n = 29)	18	50	100	100	55	69
ThinPrep (n = 60)	38	51	100	100	45	65
ThinPrep + cell block (n = 41)	26	61	80	91	40	66
Stent status (n)						
Stent present (n = 43)	17	47	100	100	74	79
No stent present (n = 210)	83	59	96	97	54	71
Type of carcinoma (n)						
Bile duct carcinoma (n = 41)	30.4					63
PDAC (n = 87)	64.4					59
Other (n = 7)	5.2					57

Abbreviations: NPV, negative predictive value; PDAC, pancreatic ductal adenocarcinoma; PPV, positive predictive value.

[a] Includes atypical, not otherwise specified + atypical, cannot rule out carcinoma.

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2017-0150-OA by guest on 07 January 2022

Table 3. Type of Carcinoma and Cytologic Diagnosis[a]			
Cytologic Diagnostic Category	Bile Duct Carcinoma, No. (%) (n = 41)	PDAC, No. (%) (n = 87)	Other, No. (%) (n = 7)
Benign, atypical, favor reactive, nondiagnostic	15 (37)	36 (41)	3 (43)
Atypical, NOS and atypical, cannot rule out carcinoma	9 (22)	21 (24)	1 (14)
Suspicious	8 (20)	12 (14)	1 (14)
Malignant	9 (22)	18 (21)	2 (29)

Abbreviations: NOS, not otherwise specified; PDAC, pancreatic ductal adenocarcinoma.

[a] There was no statistically significant difference in cytologic classification based on all 3 groups/types of carcinoma ($P = .97$) and with the "other" category excluded ($P = .84$).

with cytologic diagnoses of negative, atypical, suspicious (for malignancy), and malignant categories both individually and for the latter 3 categories as a combined group. For both these calculations, the "atypical, favor reactive" subcategory was included in the "negative" group.

Bile duct brushings with a "malignant" diagnosis had an independent sensitivity of 35%, specificity and PPV of 100%, and an accuracy rate of 66%. For specimens with a "suspicious" cytologic diagnosis, the independent sensitivity, specificity, PPV, and accuracy rates were 24%, 95%, 95%, and 63%, respectively. Similarly, specimens with an "atypical" cytologic diagnosis had an independent sensitivity of 35%, specificity of 98%, PPV of 95%, and accuracy of 65%. The NPV was 58% for each of these groups (Table 2). Absolute and relative risks of malignancy for benign, atypical, suspicious, and malignant diagnoses were also calculated. Diagnostic categories of "atypical," "suspicious," and "malignant" showed a statistically significant increase in risk of malignancy as compared to the diagnostic category of "negative" ($P < .001$) (Table 4). When 7 reviewers reviewed 30 benign and 30 malignant cases, their mean sensitivity was 71% (range, 60%–83%), mean specificity was 82% (range, 66%–93%), and accuracy was 76% (73%–80%). These results were independent of the reviewers' level of experience.[19]

Effect of Cytologic Preparation Method on Test Performance

Among the 253 patients, 46 (18%) had smears only; cytospin plus smears were made for 46 (18%), whereas 96 (38%) had ThinPrep only, and 65 (26%) had ThinPrep plus cell block available. Overall test accuracy was highest in smear preparations (83%) compared to cytospins and ThinPreps, but lower in smears-cytospin (69%), ThinPrep–cell block (66%), and ThinPrep-only (65%) combinations. Although smears had the highest accuracy of all preparation methods (except for cell block only, where there was only 1 sample), there was no statistically significant difference between smears versus ThinPrep only ($P = .08$), smears versus cytospin ($P = .21$), or smears versus ThinPrep and cell block ($P = .11$) (Table 2).

Effect of Stenting on Test Performance

Bile duct brushing sensitivity, specificity, PPV, NPV, and accuracy for patients with stents were 47%, 100%, 100%, 74%, and 79%, respectively; whereas for patients without stents, they were 59%, 96%, 97%, 54%, and 71%, respectively.

Effect of Time of Performance on Test Performance Characteristics

Time of performance of BDB as to first 7.5 years versus the second 7.5 years included in this study did not affect the accuracy of diagnosis ($P = .13$).

Effect of Type of Carcinoma on Test Performance Characteristics

Type of carcinoma (pancreatic ductal adenocarcinoma [PDAC] versus bile duct adenocarcinoma) had no statistically significant effect on accuracy ($P = .84$). Thirty-seven percent (n = 15 of 41) of bile duct carcinomas and 41% (n = 36 of 87) of PDACs were called "negative for malignant cells" or "atypical, favor reactive"; 17 of 41 bile duct carcinomas (41%) and 30 of 87 PDACs (35%) were called "suspicious" or "malignant" on cytology (Table 3).

DISCUSSION

In our cohort BDBs with a cytologic diagnosis of malignant had an overall low sensitivity (35%) and NPV (58%) in diagnosing malignancy with contrasting high specificity and PPV (both 100%), as well as moderate accuracy (66%). In other words, while only one-third of cancers were correctly identified by this method, the specificity and PPV showed that there were no false-positive results during the 15-year period. While commendable, these excellent specificity results were obtained at the expense of sensitivity, a finding that highlights cytopathologists' reluctance to call brushings malignant unless the cytologic evidence is qualitatively and quantitatively overwhelmingly convincing (and most likely high-grade/poorly differentiated tumors). An alternative explanation is that low sensitivity may be due to procedural undersampling rather than interpretative reluctance, but the data on our "suspi-

Table 4. Overall Risk of Malignant Outcome With Respect to Bile Duct Brushing Result in Patients With Follow-up (n = 253)						
Cytology Diagnosis	Benign Outcome	Malignant Outcome	% Malignant	Relative Risk	95% CI	*P* Value
Benign (n = 159)[a]	93	66	42%	1.00	...	...
Atypical (n = 37)[b]	2	35	95%	2.28	1.87–2.78	<.001
Suspicious (n = 22)	1	21	96%	2.30	1.87–2.83	<.001
Malignant (n = 35)	0	35	100%	2.41	2.00–2.90	<.001

[a] Includes nondiagnostic, benign, and atypical, favor reactive.
[b] Includes "atypical (not otherwise specified)" and "atypical, cannot rule out carcinoma" categories.

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2017-0150-OA by guest on 07 January 2022

cious" and "atypical" categories would suggest otherwise, since the RRs of carcinoma for these categories were very similar to that of the "malignant" category, indicating that these specimens contain findings that are probably representative of malignancy but cannot currently be recognized as such (see more on this below). Although low specimen cellularity and/or poor preservation can play a major role in interpretation, this did not appear to be the case in our cohort, as only 2.4% of cases were deemed nondiagnostic. Nonetheless, it is noteworthy that mechanical issues, such as skill of the endoscopist in cannulation, placement and brushing of the bile duct, as well as the tumor cells' potential to exfoliate, also play a critical role in the quality of the samples obtained for interpretation.

Our test sensitivity rate is at the lower end of the spectrum of some previously reported in the literature (range, 30%–68%),[14–16] but it is higher than rates reported by others (6%–33%),[3–5,7,8,22] which could be the result of several factors, some of which we explored in detail.

During the 15-year period our BDB specimens were reviewed by 11 or more fellowship-trained cytopathologists, each with variable degrees of experience. Pathologist inexperience could certainly have played a role in low test sensitivity. However, because of the large number of reviewers during the period studied, it was hard to collect meaningful data regarding each individual pathologist's performance. Interpretative errors are certainly a known problem with BDBs, as evidenced by studies highlighting low interobserver reproducibility of cytologic categories[17] and even high intraobserver variations in diagnoses.[3] However, our recent study evaluating interreviewer agreement in evaluating BDBs found that both experienced and inexperienced reviewers had similar performance with respect to sensitivity, specificity, and overall accuracy.[19] Additionally, the protocol for obtaining and analyzing BDBs at our institution did not change between 2000 and 2015. The specificity rate, however, was very high in this study (100%). Previous studies also report high specificities, ranging between 87% and 100%.[14] It appears that this reflects an overall tendency in the cytopathology community to acknowledge malignancy only when glaringly evident, and suggests an overall reluctance to render a diagnosis of malignancy outside of this setting (as illustrated by the similar RR of malignancy among the "suspicious/atypical" groups compared to that of the "malignant" category in our cohort). This reluctance is also reflected by the high frequency of ambiguous diagnoses in our cohort, with 35% of cases with adequate follow-up having ambiguous diagnoses (26% atypical and 9% suspicious). Moreover, when "atypical," "suspicious," and "malignant" diagnostic categories were combined, our test's sensitivity almost doubled to 58%, but led to only a slight decline in specificity and PPV figures, which both dropped minimally to 97%, while overall accuracy increased to 73%.

Our observations are consistent with previous studies by Volmar et al[18] and Chadwick et al.[14] "Atypical" and "suspicious for malignancy" diagnoses are rendered when the specimen demonstrates worrisome cytologic features such as abnormal architectural and nuclear changes that are insufficient for a definitive diagnosis of malignancy, generally because of quantitative or qualitative issues (scant cellularity or degenerative changes being most frequent). In our series, for patients with confirmed malignancy, the "atypical (NOS and cannot exclude carcinoma)" and "suspicious" cases both had similar RR of malignancy of 2.3, which was very close to that of the "malignant" group (RR, 2.41) (Table 3). In fact, most (≥95%) "atypical" (NOS and cannot exclude carcinoma) and "suspicious" specimens proved to have a malignant outcome. Thus, clinical management teams ought to keep in mind that the "atypical" and "suspicious" diagnostic categories should also be considered high risk, and such patients should have early repeated brushings with hopefully improved specimen quality along with submission of a sample for ancillary testing (particularly fluorescence in situ hybridization) and biopsy, if possible.

We investigated the potential influence of various parameters on the interpretation of BDBs. Of all preparation methods used at our institution, direct smears had the highest sensitivity and accuracy both when used alone and in combination with cytospin, while ThinPrep had one of the lowest sensitivity rates and the lowest accuracy rate for detecting malignancy. In fact, direct smears trended toward higher (though not statistically significant) accuracy rates than ThinPreps ($P = .08$). These trends may presumably be related to cytopathologists' greater familiarity and comfort with smears than ThinPrep and other methods. The ThinPrep system was approved for use in gynecologic cytology in 1996 and since then it has been an alternative preparation method for nongynecologic specimens including thyroid, breast, and biliary tract. In a study by Ylagan et al,[23] ThinPrep had 3 advantages: (1) elimination of air-drying artifact, (2) elimination of blood-clotting artifact, and (3) decreasing cell overlap. However, there was also a tendency for cells to form micropapillary-like clusters with rounded edges, despite benignity and a propensity for cell shrinkage and loss of necrotic debris, which can make these samples more likely to be misinterpreted as malignant, especially in comparison to smears. Additionally, the clustered cell groups in liquid-based preparations absorb more light and thus become darker, which obscures their cytologic features. Some studies have shown a statistically significant increase in BDB sensitivity and diagnostic accuracy when smears and ThinPrep are combined, and have found the combination to be superior to either direct smear with cytospin or direct smear alone.[18] Unfortunately, at our institution, we do not use this particular combination of preparations for BDB but it warrants investigation. In contrast, Okonkwo et al[24] showed that cytocentrifuged preparations made from brush tip material were superior to direct smears for evaluating bile duct epithelium. Interestingly, when we compared specimens with ThinPrep alone versus those with both ThinPrep and cell block, the latter had higher sensitivity (61%) and lower specificity (80%) (versus 51% and 100%, respectively). Thus, it would appear that cell blocks are for the most part able to capture more malignant cases, but at the risk of overinterpretation of benign and/or reactive changes as malignant, which suggests that while they are aiding in the accurate diagnosis of malignancy in many cases, they are also prone to overinterpretation, and should always be evaluated with caution.

Pancreatic adenocarcinomas (64.5%) and bile duct carcinomas (30.4%) constituted the vast majority of tumors detected on BDBs in our cohort and had a 2:1 ratio. However, the type of carcinoma had no significant effect on test accuracy ($P = .60$) or the frequency of any of the diagnostic categories. These observations are similar to those of Choi et al,[25] Volmar et al,[18] and Okonkwo et al.[24] One might expect that bile duct carcinoma, which can be

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2017-0150-OA by guest on 07 January 2022

extremely well differentiated, would cause more diagnostic difficulty than PDAC; however, this was not the case in our study. In a study by Okonkwo et al,[24] all 13 PDACs were diagnosed as malignant in BDB, whereas only 4 of 7 cholangiocarcinomas were accurately diagnosed as malignant. However, Adamsen et al[3] reported highest sensitivity for bile duct carcinoma, intermediate sensitivity for PDAC, and lowest sensitivity in biliary obstruction due to metastatic disease. These differences may be related to case selection bias. Similarly, Barr Fritcher et al[26] found that the ratio of PDAC to cholangiocarcinoma was reversed, with 60.7% bile duct carcinomas and 35.7% PDACs, which may be population or case selection related. The cohort of Barr Fritcher and colleagues[26] included 1 primary gallbladder carcinoma and 1 ampullary adenocarcinoma among positive cases. In our study, 3 gallbladder carcinomas, 2 neuroendocrine tumors, 1 hepatocellular carcinoma, and 1 metastatic colonic adenocarcinoma were identified among the malignant cases, with the hepatocellular carcinoma being initially interpreted as an adenocarcinoma. Choi et al[25] found 6 ampullary adenocarcinomas, 3 gallbladder carcinomas, and 1 neuroendocrine neoplasm in their series. Clearly, a wide variety of tumors can primarily or secondarily involve the biliary tract, including metastatic colonic, lung small cell, and esophageal carcinoma.[18,24] Interestingly, 1 metastatic carcinoma and 3 biliary tract lymphomas were initially missed on BDB.[18,24] These studies highlight BDB's poor sensitivity in diagnosing metastatic carcinoma, lymphoma, as well as hepatocellular carcinoma, a factor previously reported by others.[27,28]

Another interesting (but perhaps not surprising) finding was the fact that sensitivity was higher in nonstented specimens than those with stents (59 versus 47%), while specificity was lower in nonstented BDBs relative to stented ones (96 versus 100%). The higher specificity, PPV, NPV, and accuracy but lower sensitivity seen in our stented versus nonstented cases suggests that cytopathologists are often reluctant to render a malignant diagnosis in stented specimens (with the legitimate concern that the atypia may be reactive in nature), unless the cytologic and/or clinical evidence is overwhelmingly convincing. Biliary tract instrumentation can cause architectural distortion of biliary epithelium as well as nuclear atypia, which can easily mimic malignancy on cytology. While our results show some effect of stenting on cytopathologists' interpretation of BDBs, previous reports suggest that prior stenting did not affect the diagnostic value of BDB.[29] Volmar et al[18] also showed that the presence of a stent did not have a significant effect on test performance.

CONCLUSIONS

While BDB sensitivity was fairly low for malignancy (with only one-third of cancers being correctly identified), specificity and PPV were high with no false positives. When atypical, suspicious, and malignant diagnostic categories were combined, specificity and PPV dropped minimally but sensitivity almost doubled. Presumably because of operator familiarity, smears trended toward higher (though not significant) accuracy than ThinPreps. Cell blocks are helpful in clarifying malignancy but should be interpreted with caution because of a risk of overdiagnosis of malignancy with this preparation method. Type of malignancy (bile duct versus PDAC) had no effect on accuracy rates. The higher specificity, PPV, NPV, and accuracy but lower sensitivity rates seen for our patients with stents versus without stents suggest that such cases are only diagnosed as malignant when the evidence is overwhelmingly convincing. Although the Papanicolaou Society recently published its reporting system for pancreatobiliary cytology, this report placed emphasis on guidelines for terminology and nomenclature with less emphasis on cytomorphology, particularly in the evaluation of biliary tract pathology.[30] Therefore, more in-depth studies are needed to examine the most useful cytomorphologic and/or molecular features and criteria that can best identify malignancy and distinguish it from its benign and reactive mimics, since this anatomic site is especially prone to inflammatory and instrumentation-related alteration.

The field of biliary tract cytopathology is still going through a "learning curve." As these specimens become more commonplace and as both the operators and interpreters become more experienced with them, cytologic diagnostic criteria will likely improve, as it has for other sites including salivary gland and urine. However, the innate challenges presented by pancreatobiliary cancers, including their close mimicry of reactive changes, paucicellularity of these carcinomas (because of abundant stroma), and difficulties in acquiring adequate samples, will likely make BDB interpretation an ongoing challenge. Fluorescence in situ hyridization and molecular testing show great promise as valuable complementary adjuncts in the evaluation of these specimens.

References

1. Layfield LJ, Wax TD, Lee JG, Cotton PB. Accuracy and morphologic aspects of pancreatic and biliary duct brushings. *Acta Cytol*. 1995;39(1):11–18.
2. Logrono R, Kurtycz D, Molina C, Trivedi V, Wong J, Block K. Analysis of false-negative diagnoses on endoscopic brush cytology of biliary and pancreatic duct strictures. *Arch Pathol Lab Med*. 2000;124(3):387–392.
3. Adamsen S, Olsen M, Jendresen MB, Holck S, Glenthoj A. Endobiliary brush biopsy: intra- and interobserver variation in cytological evaluation of brushings from bile duct strictures. *Scand J Gastroenterol*. 2006;41(5):597–603.
4. Burnett AS, Calvert TJ, Chokshi RJ. Sensitivity of endoscopic retrograde cholangiopancreatography standard cytology: 10-y review of the literature. *J Surg Res*. 2013;184(1):304–311.
5. Draganov PV, Chauhan S, Wagh MS, et al. Diagnostic accuracy of conventional and cholangioscopy-guided sampling of indeterminate biliary lesions at the time of ERCP: a prospective, long-term follow-up study. *Gastrointest Endosc*. 2012;75(2):347–353.
6. Duggan MA, Brasher P, Medlicott SA. ERCP-directed brush cytology prepared by the Thinprep method: test performance and morphology of 149 cases. *Cytopathology*. 2004;15(2):80–86.
7. Govil H, Reddy V, Kluskens L, et al. Brush cytology of the biliary tract: retrospective study of 278 cases with histopathologic correlation. *Diagn Cytopathol*. 2002;26(5):273–277.
8. Hart J, Parab M, Mandich D, Cartun RW, Ligato S. IMP3 immunocytochemical staining increases sensitivity in the routine cytologic evaluation of biliary brush specimens. *Diagn Cytopathol*. 2012;40(4):321–326.
9. Khalid A, Pal R, Sasatomi E, et al. Use of microsatellite marker loss of heterozygosity in accurate diagnosis of pancreaticobiliary malignancy from brush cytology samples. *Gut*. 2004;53(12):1860–1865.
10. Mohammad Alizadeh AH, Mousavi M, Salehi B, et al. Biliary brush cytology in the assessment of biliary strictures at a tertiary center in Iran. *Asian Pac J Cancer Prev*. 2011;12(10):2793–2796.
11. Papaxoinis K, Patsouris E, Athanassiadou P, Nicolopoulou-Stamati P. Contribution of nuclear morphometry by confocal laser scanning microscopy to the diagnosis of malignant bile duct strictures. *Acta Cytol*. 2009;53(2):137–143.
12. Stoos-Veic T, Bilic B, Kaic G, Ostovic KT, Babic Z, Kujundzic M. Biliary brush cytology for the diagnosis of malignancy: a single center experience. *Coll Antropol*. 2010;34(1):139–143.
13. Cohen MB, Wittchow RJ, Johlin FC, Bottles K, Raab SS. Brush cytology of the extrahepatic biliary tract: comparison of cytologic features of adenocarcinoma and benign biliary strictures. *Mod Pathol*. 1995;8(5):498–502.
14. Chadwick BE, Layfield LJ, Witt BL, Schmidt RL, Cox RN, Adler DG. Significance of atypia in pancreatic and bile duct brushings: follow-up analysis of the categories atypical and suspicious for malignancy. *Diagn Cytopathol*. 2014;42(4):285–291.
15. Salomao M, Gonda TA, Margolskee E, et al. Strategies for improving diagnostic accuracy of biliary strictures. *Cancer Cytopathol*. 2015;123(4):244–252.

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2017-0150-OA by guest on 07 January 2022

16. Tamada K, Ushio J, Sugano K. Endoscopic diagnosis of extrahepatic bile duct carcinoma: advances and current limitations. *World J Clin Oncol*. 2011;2(5):203–216.
17. Layfield LJ, Schmidt RL, Chadwick BE, Esebua M, Witt BL. Interobserver reproducibility and agreement with original diagnosis in the categories "atypical" and "suspicious for malignancy" for bile and pancreatic duct brushings. *Diagn Cytopathol*. 2015;43(10):797–801.
18. Volmar KE, Vollmer RT, Routbort MJ, Creager AJ. Pancreatic and bile duct brushing cytology in 1000 cases: review of findings and comparison of preparation methods. *Cancer*. 2006;108(4):231–238.
19. Avadhani V, Hacihasanoglu E, Memis B, et al. Cytologic predictors of malignancy in bile duct brushings: a multi-reviewer analysis of 60 cases. *Mod Pathol*. 2017;30(9):1273–1286.
20. Gonzalez RS, Bagci P, Basturk O, et al. Intrapancreatic distal common bile duct carcinoma: analysis, staging considerations, and comparison with pancreatic ductal and ampullary adenocarcinomas. *Mod Pathol*. 2016;29(11):1358–1369.
21. American Cancer Society. Bile Duct Cancer (Cholangiocarcinoma). In: American Cancer Society, ed. 2017. https://old.cancer.org/acs/groups/cid/documents/webcontent/003084-pdf.pdf; 2016. Accessed October 4, 2017.
22. Kipp BR, Stadheim LM, Halling SA, et al. A comparison of routine cytology and fluorescence in situ hybridization for the detection of malignant bile duct strictures. *Am J Gastroenterol*. 2004;99(9):1675–1681.
23. Ylagan LR, Liu LH, Maluf HM. Endoscopic bile duct brushing of malignant pancreatic biliary strictures: retrospective study with comparison of conventional smear and ThinPrep techniques. *Diagn Cytopathol*. 2003;28(4):196–204.
24. Okonkwo AM, De Frias DV, Gunn R, et al. Reclassification of "atypical" diagnoses in endoscopic retrograde cholangiopancreaticography-guided biliary brushings. *Acta Cytol*. 2003;47(3):435–442.
25. Choi WT, Swanson PE, Grieco VS, Wang D, Westerhoff M. The outcomes of "atypical" and "suspicious" bile duct brushings in the identification of pancreaticobiliary tumors: follow-up analysis of surgical resection specimens. *Diagn Cytopathol*. 2015;43(11):885–891.
26. Barr Fritcher EG, Caudill JL, Blue JE, et al. Identification of malignant cytologic criteria in pancreatobiliary brushings with corresponding positive fluorescence in situ hybridization results. *Am J Clin Pathol*. 2011;136(3):442–449.
27. Rupp M, Hawthorne CM, Ehya H. Brushing cytology in biliary tract obstruction. *Acta Cytol*. 1990;34(2):221–226.
28. Vadmal MS, Byrne-Semmelmeier S, Smilari TF, Hajdu SI. Biliary tract brush cytology. *Acta Cytol*. 2000;44(4):533–538.
29. Ponsioen CY, Vrouenraets SM, van Milligen de Wit AW, et al. Value of brush cytology for dominant strictures in primary sclerosing cholangitis. *Endoscopy*. 1999;31(4):305–309.
30. Pitman MB, Centeno BA, Ali SZ, et al. Standardized terminology and nomenclature for pancreatobiliary cytology: the Papanicolaou Society of Cytopathology guidelines. *Diagn Cytopathol*. 2014;42(4):338–350.

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2017-0150-OA by guest on 07 January 2022