

Cytologic predictors of malignancy in bile duct brushings: a multi-reviewer analysis of 60 cases

Vaidehi Avadhani¹, Ezgi Hacıhasanoğlu¹, Bahar Memis¹, Burcin Pehlivanoglu¹, Krisztina Z Hanley¹, Uma Krishnamurti¹, Alyssa M Krasinskas¹, Adeboye O Osunkoya¹, Lauren M Daniels², Alexa A Freedman², Michael Goodman², Volkan Adsay¹ and Michelle D Reid¹

¹Department of Pathology, Emory University School of Medicine, Atlanta, GA, USA and ²Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA, USA

Diagnosing malignancy in bile duct brushings is highly challenging. Seven reviewers of variable backgrounds and levels of participation in bile duct brushing sign out blindly reviewed 60 specimens (30 malignant with histologic confirmation and 30 benign (15 stented) with resection or ≥ 18 months of uneventful follow-up), testing the utility of 14 malignant characteristics. Eleven characteristics were statistically significantly associated with malignancy including 3-dimensional clusters (63% in malignant vs 3% in benign, odds ratio 50, $P=0.0003$), pleomorphism (62 vs 3, odds ratio 48, $P=0.0004$), 2-cell population (60% vs 3, odds ratio 44, $P=0.0005$), chromatin pattern (hypo/hyperchromasia) changes (70% vs 7%, odds ratio 33, $P<0.0001$), high nuclear-to-cytoplasmic ratio (48 vs 3%, odds ratio 27, $P=0.0023$), cytoplasmic vacuoles (43 vs 3%, odds ratio 22, $P=0.0042$), nuclear irregularity (70 vs 10%, odds ratio 21, $P<0.0001$), cellular discohesion (38 vs 3%, odds ratio 18, $P=0.0082$), hypercellularity (23% vs 0), nuclear molding (20% vs 0) and prominent nucleoli (21% vs 0). Necrosis and infiltrating inflammation were not helpful in identifying malignancy ('neutrophil cannibalism' was noted in 43% malignant); 21/30 (70%) malignant brushings had ≥ 3 malignant characteristics, while 23 (77%) benign brushings had none. Of 20 brushings with ≥ 4 characteristics, 1(5%) proved benign and showed detachment atypia, a close malignant mimicker in brushings. Identification of 3 characteristics maximized the combined sensitivity (70%), specificity (97%) and accuracy (83%), but sensitivity dropped as number of characteristics increased. Identification of 3/11 characteristics (3-dimensional clusters, pleomorphism, high nuclear-to-cytoplasmic ratio, nuclear irregularity, hypercellularity, discohesion, chromatin changes, vacuoles, prominent nucleoli, molding and 2-cell population) improves pathologists' overall performance greatly.

Modern Pathology (2017) 30, 1273–1286; doi:10.1038/modpathol.2017.51; published online 30 June 2017

Bile duct brushings are often used as the initial investigative pathology test for pancreatobiliary tract lesions as they widely sample the bile duct and have a low complication rate. Unfortunately, bile duct brushings are some of the most challenging cytologic specimens to evaluate due in part to poor specimen quality and quantity, as well as frequent ulceration, inflammation, and stent-related atypia that may make the distinction of benign or reactive biliary

epithelium from neoplasms particularly challenging.¹ Additionally, the high frequency of deceptively benign-appearing carcinomas in this site accounts for a significant number of false-negative results.² As a result of these factors cytological diagnosis of malignancy on bile duct brushing is notoriously limited with low sensitivity rates (6–64%; mean 42%) and negative predictive value,^{3–12} despite high test specificity (98–100%).^{2,13,14}

Because of newly improved imaging modalities bile duct brushings are steadily increasing and cytologic abnormalities are being identified in these samples even before a mass is visible. Early diagnosis of pancreatobiliary tract carcinoma ensures early treatment while for unresectable tumors accurate diagnosis preempts unnecessary

Correspondence: Dr MD Reid, MD, Department of Pathology, Emory University Hospital, Department of Pathology and Laboratory Medicine, 1364 Clifton Road NE, Room H-189, Atlanta, GA 30322, USA.

E-mail: michelle.reid@emory.edu

Received 13 December 2016; revised 15 March 2017; accepted 19 April 2017; published online 30 June 2017

surgery and ensures appropriate chemo-radiation and/or palliative therapy.

Several studies have attempted to identify definite cytologic criteria that can better predict malignancy in bile duct brushings.^{1,5,14–17} In 1995, Cohen *et al* were among the first to use cytologic criteria to improve diagnostic accuracy.¹³ Their analysis showed that nuclear molding, chromatin clumping and increased nuclear-cytoplasmic ratio were frequently associated with malignancy (deemed ‘primary Iowa criteria’). If all primary criteria were not present, secondary criteria (anisonucleosis, nuclear irregularity/grooves, and nuclear enlargement) were used. Presence of 2/3 primary criteria resulted in 83% sensitivity and 98% specificity for carcinoma detection. In 1998 Renshaw *et al* similarly showed that nuclear molding, chromatin clumping and loss of polarity were associated with malignancy (deemed the ‘Boston criteria’).¹⁷ Henke *et al* were later successfully able to apply the Iowa criteria to liquid-based specimens.¹⁸ Others have since examined different cytologic criteria to diagnose cholangiocarcinoma on bile duct brushings, with variable success.^{1–9} Barr Fritcher *et al* recently examined 16 cytologic criteria associated with malignancy in pancreatobiliary brushings with corresponding positive fluorescence *in situ* hybridization (FISH), and found that abnormal single cells, nuclear membrane irregularity and enlargement were independently associated with malignancy.¹⁴

Most major previous studies typically involved 2–4 reviewers,^{14,17,19,20} some of them cytopathology experts, thus their observations may not be translatable to daily practice where bile duct brushings are frequently evaluated by general surgical pathologists with limited cytopathology or gastrointestinal pathology experience. Additionally, in some studies, patients being alive after 6 months was considered evidence of benignity of the lesion,²¹ although findings in more recent observations in bile duct carcinoma suggest that this is too short a time frame in which to exclude malignancy.^{22,23}

In this study, we investigated the specific cytologic criteria that lead pathologists from various backgrounds (but who are involved in the sign out of these specimens) to the accurate diagnosis of malignancy in 60 BDB samples with >18 months follow-up or definitive tissue diagnosis of malignancy.

Materials and methods

Case Selection

After approval by our institution’s Review Board, a computer-based search of the Emory University Pathology Department’s archives 444 bile duct brushings were collected over a 16-year period (2000–2015). Among the 444 bile duct brushings available for review in our Institution, 253 fulfilled the follow-up criteria (see below) and from these 30

malignant and 30 benign samples were randomly selected by a non-reviewing author (EH) for inclusion in the study. After consultation with our statisticians a sample size of 60 was determined based on the ability to detect a proportion (sensitivity or specificity) of 80% with a 95% confidence interval from 70–90%.²⁴ All 30 malignant cases had a definitive histologic diagnosis of adenocarcinoma either by biopsy ($n=6$) or resection ($n=24$) and was verified by the authors. All 30 benign cases had 18-plus months of uneventful follow-up, 15 of them with prior stent placement and 15 without. Patient demographics, imaging details, cytologic and histologic diagnoses, and clinical follow-up were collected. Follow-up histologic biopsies, resections and other FNAs of primary and/or metastatic tumors were reviewed in all cases.

All bile duct brushings had been obtained during endoscopic retrograde cholangiopancreatography (ERCP). These were directly smeared and stained with Papanicolaou stain or collected in ethanol or liquid-based cytology preservative and ThinPrep smears were prepared and stained with Papanicolaou stain. Where possible, cell blocks were also prepared from paraffin embedded material.

Cytopathologic Criteria

Based on the results of prior studies and the authors’ personal pathology experience,^{1–5,9} the presence or absence of 14 malignant characteristics which included (1) hypercellularity, (2) 3-dimensional clusters, (3) cellular discohesion characterized by single cells with high (>50%) nuclear to cytoplasmic ratio, (4) high nuclear to cytoplasmic ratio (>50%), (5) nuclear molding or hugging, (6) cytoplasmic mucin vacuoles, (7) 2-cell population, (8) nuclear chromatin changes (hypochromasia or hyperchromasia), (9) nuclear membrane irregularity, (10) large prominent nucleoli, (11) nuclear pleomorphism, (12) necrosis (single cell or background), (13) abnormal mitoses, and (14) infiltrating inflammation within epithelial cells (Figures 1, 2, 3, 4, 5, 6, and 7). High nuclear to cytoplasmic ratio (>50%) in malignant cells was defined as a nuclear to cytoplasmic ratio that was greater than 2:1 or in which the nucleus constituted more than 50% of the entire cell volume.

Review

Cytology slides from the 60 bile duct brushings were de-identified and given unique identification numbers known only to one study participant (EH) who was not a reviewer in the study. A log form with 14 established malignant cytologic criteria was designed to facilitate the 7 blinded reviewers’ entry of their cytologic diagnoses. The reviewers were instructed to render a diagnosis of benign vs malignant for each case, and to document the presence or absence of the 14 criteria in each case.

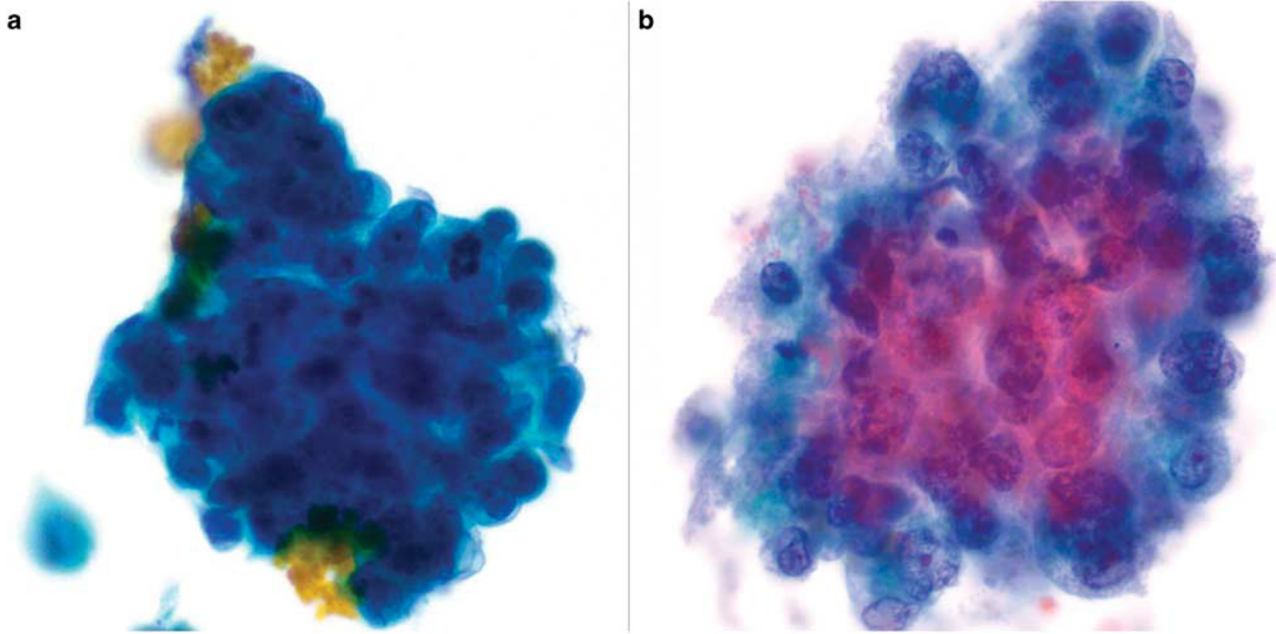


Figure 1 (a,b) Three-dimensional clusters of malignant cells with high nuclear to cytoplasmic ratio, hyperchromasia, nuclear contour irregularity and prominent nucleoli. Note the clinging tumor diathesis in (b) (Papanicolaou stain, magnification $\times 200$).

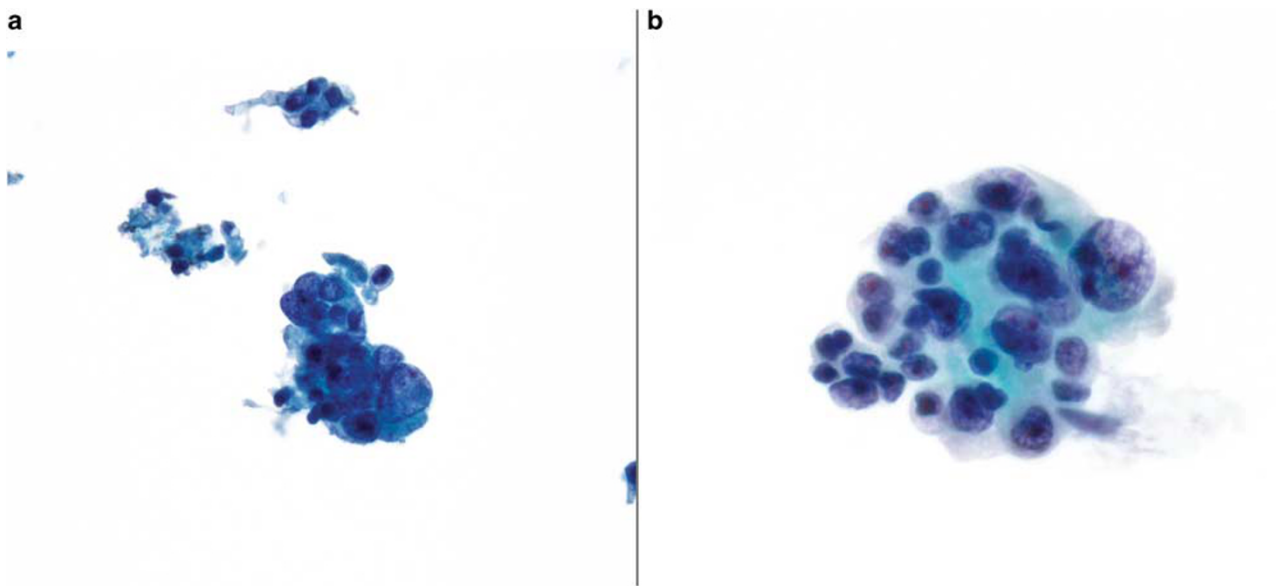


Figure 2 (a,b) Clusters of malignant cells with marked nuclear pleomorphism, 4-fold anisonucleosis, high nuclear to cytoplasmic ratio and hyperchromasia (Papanicolaou stain, magnification $\times 200$ (a) and $\times 400$ (b)).

The 7 reviewers included 3 fellowship-trained cytopathologists (MDR, KH,UK- reviewers 1, 2, and 6), 2 surgical pathologists with pancreatobiliary pathology expertise (one gastrointestinal pathology fellowship-trained with limited cytology experience (AK; reviewer 3) and the other an oncology fellowship-trained pathologist with expertise in pancreatobiliary pathology who routinely signed out pancreatobiliary cytology (VoA; reviewer 4)), 1 general surgical and genitourinary pathology-trained pathologist with sign out responsibility in cytopathology (AOC; reviewer 5) and 1 cytopathology

fellow with 3 months of training (VaA; reviewer 7). Reviewers were blinded to all clinical or radiologic information, histologic diagnoses, patient outcome, as well as other reviewers' diagnoses. All 60 BDB samples were then blindly reviewed, corresponding log forms were completed and ultimately tabulated.

Statistical Analysis

The utilization of the 14 characteristics in the accurate cytologic diagnosis by the 7 reviewers was

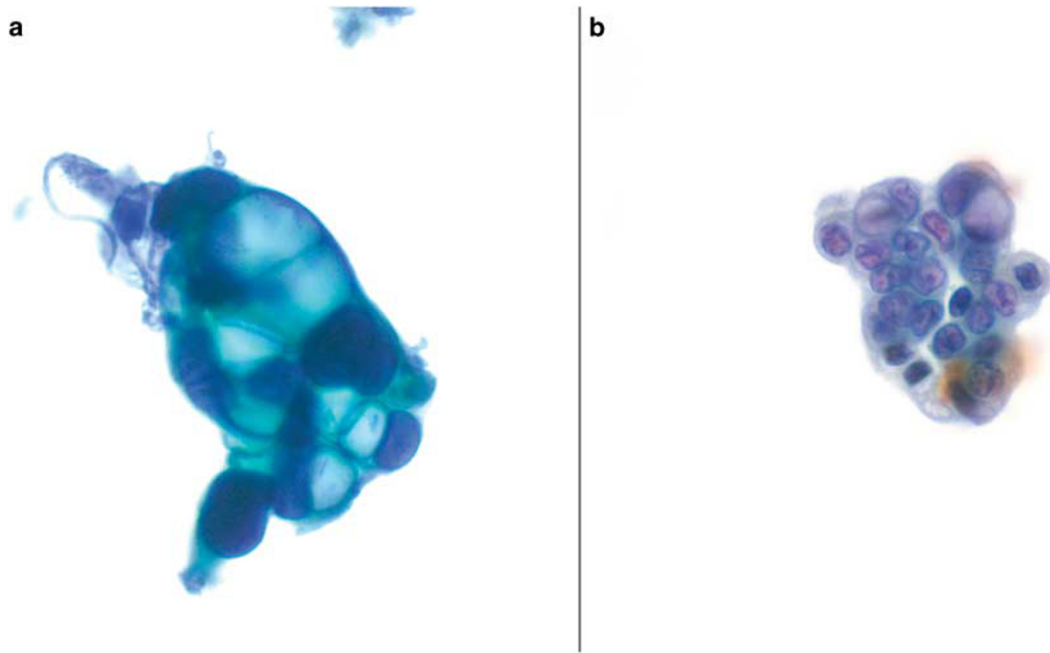


Figure 3 (a,b) These malignant cells range from hyperchromatic (**a**) to hypochromatic (**b**) and have large circumscribed intracytoplasmic mucin vacuoles that focally displace their nuclei eccentrically (Papanicolaou stain, magnification $\times 200$).

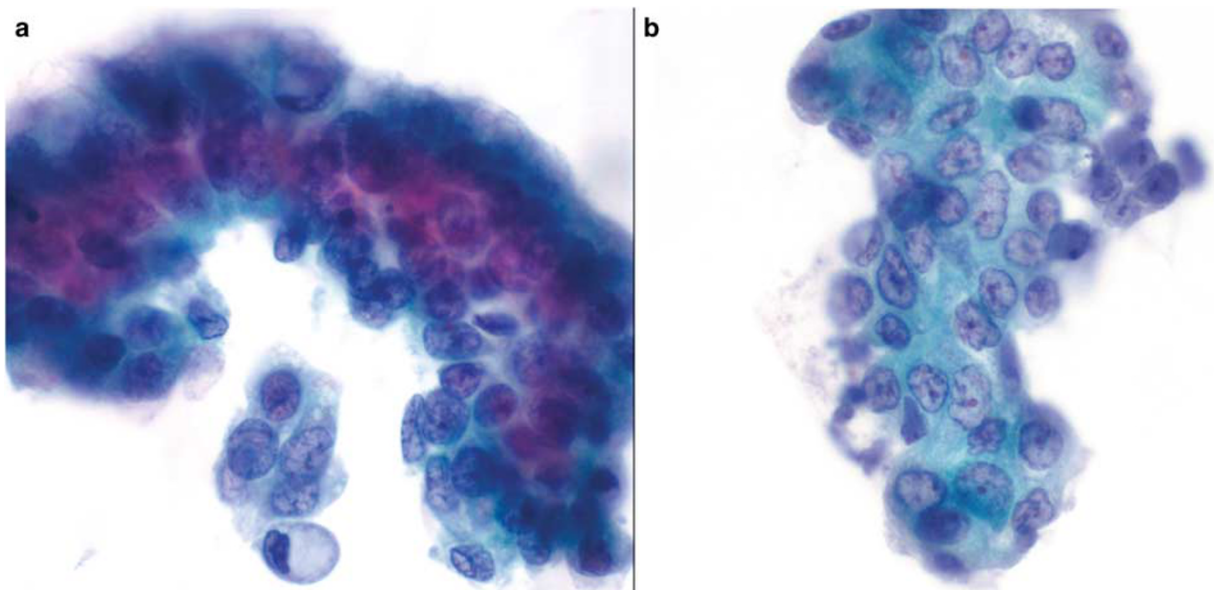


Figure 4 (a,b) These malignant cells are crowded and overlapping with high nuclear to cytoplasmic ratio, marked nuclear contour irregularity and change in chromatin pattern, with hyperchromasia (in **a**) and hypochromasia (in **b**) as well as focal cellular dissociation and intracytoplasmic mucin vacuoles (in **a**). Prominent red nucleoli are also present in (**b**) (Papanicolaou stains, magnification $\times 400$).

analyzed through different approaches. To be able to perform the statistical analysis our expert statisticians (MG, LD, and AF) determined that a 'gold standard' would need to be developed to determine the presence or absence of the 14 characteristics. In this manuscript, we present results using three different reviewer groups as 'gold standard'. In the primary analysis, agreement by two of three cytopathologists (MDR, KH, UK) was used as the gold

standard for the parameter validation, as all three cytopathologists were board certified and had similar or comparable years of experience (10, 7, and 11 years respectively). There were two samples where one of the cytopathologists reported the sample as non-diagnostic and did not review the criteria. In these samples, agreement was based on 2/2 reviewers and if the reviewers did not agree, then the characteristic was considered inconclusive and

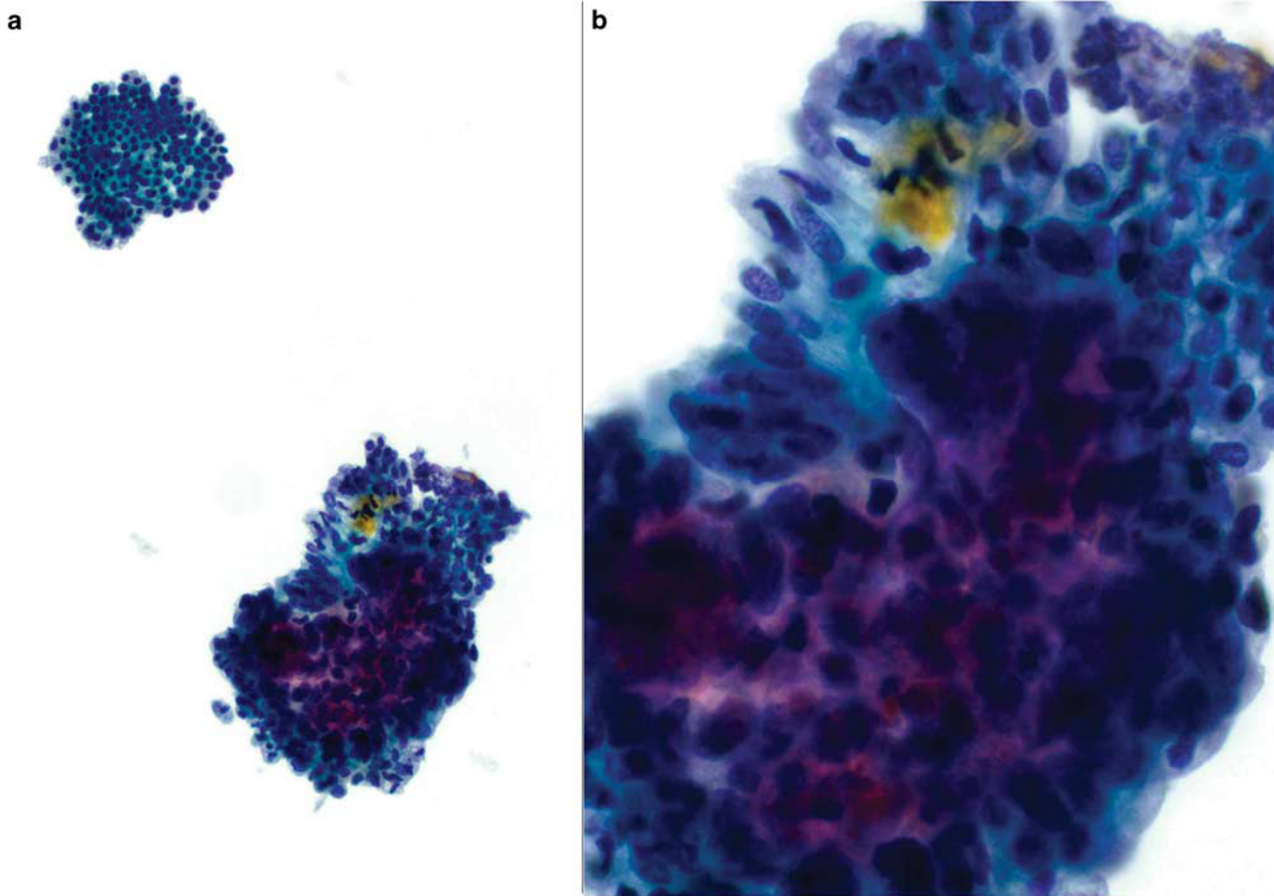


Figure 5 (a,b) This examples of a 2-cell population shows a flat honeycomb sheet of evenly distributed benign columnar cells with low nuclear to cytoplasmic ratio and even chromatin distribution next to a crowded cluster of malignant cells with marked nuclear crowding and overlapping, high nuclear to cytoplasmic ratios and hyperchromasia (Papanicolaou stains, magnification $\times 200$ (a) and magnification $\times 400$ (b)).

the sample was excluded from analysis of that particular characteristic. Separately, 'gold standards' based on agreement between any 4 of 7 reviewers (independent of cytopathology training and experience) was also determined as well as agreement between 7 of 7 reviewers. Agreement between 7 of 7 reviewers reflects the most stringent analysis because a characteristic is only considered present if all 7 reviewers identified it and absent if < 7 reviewers identified it.

A score based on the number of malignant characteristics present was calculated for each BDB sample and analyzed as a continuous variable. Sensitivity, specificity, and accuracy were evaluated for each cut point of this continuous variable to determine the cut point that maximized these proportions. Following this, once the parameters with strong association were established, then the potential value of combination of parameters were also tested by using backwards elimination approach, which was used to reduce the model containing all characteristics to the number of characteristics indicated by the best performing cut point to determine if a specific set of fewer variables could be used to determine a diagnosis. *P*-values for

differences between the malignant and benign groups were calculated using chi-square tests. Logistic regression was used to calculate odds ratios for individual characteristics and for malignancy score. All statistical tests used a *P*-value of < 0.05 to determine significance. Analyses were conducted using SAS, version 9.4 (SAS Institute, Cary, NC, USA).

Results

Cytologic Findings Using Agreement by 2/3 Cytopathologists as the Gold Standard

Of the 14 malignant characteristics, only 11 were statistically significantly associated with malignancy while two were not (inflammation ($P=0.4237$) and necrosis ($P=0.7188$)). Mitoses could not be evaluated because there was no agreement between 2 or more cytopathologists on the presence of characteristic in any of the 60 bile duct brushings (Table 1). Among all malignant cases the most frequent characteristics (from greatest to least) were change in chromatin pattern (hypo/hyperchromasia (70%), nuclear irregularity (67%), pleomorphism (62%),

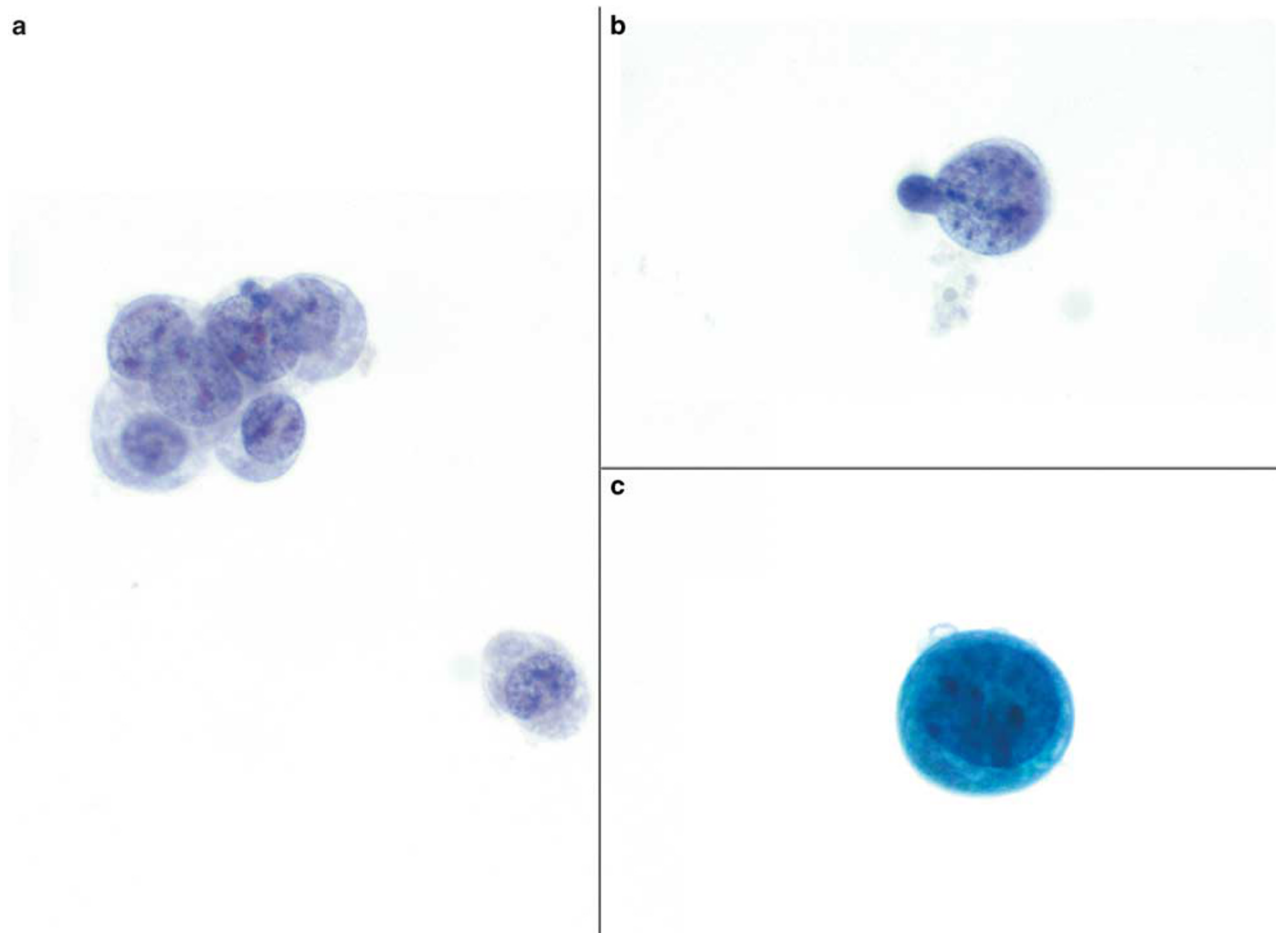


Figure 6 (a) Cellular dissociation of single intact malignant cells (Papanicolaou stain, magnification $\times 200$) with high nuclear cytoplasmic ratio ($>50\%$) (b,c) are shown in these examples (Papanicolaou stains, magnification $\times 400$).

2-cell population (57%), 3-dimensional clusters (52%), high nuclear to cytoplasmic ($>50\%$) ratio (48%), cytoplasmic mucin vacuoles (43%), inflammation (43%), cellular discohesion of cells with high N/C ratio (38%), hypercellularity (23%), prominent nucleoli (21%), and necrosis (17%).

Among benign cases, the frequency of malignant characteristics (from greatest to least) was infiltrating inflammation (33%), nuclear irregularity (13%), necrosis (13%), cytoplasmic mucin vacuoles (13%), 3-dimensional clusters (3%), cellular discohesion (3%), high N/C (3%), and pleomorphism (3%). There was no agreement between 2 or more cytopathologists on the presence of hypercellularity, nuclear molding, 2-cell population and prominent nucleoli in any of the benign cases.

Interestingly, among all 60 Bile duct brushings, 'necrosis' was recorded with similar frequency in malignant and benign cases (17% vs 13%, P -value 0.718). Inflammation was present in 33% of benign bile duct brushings and was greater in those without stents (40% vs 27%). The prevalence of malignant characteristics in all malignant and all benign Bile duct brushings as analyzed by the cytopathologists is summarized in Figure 8.

Cytologic Findings Using Agreement by 4/7 Reviewers as the Gold Standard

When agreement between 4/7 reviewers was used as the gold standard, 11/14 malignant characteristics were statistically significantly associated with a malignant diagnosis while two characteristics (inflammation ($P=0.4051$) and nuclear molding/hugging ($P=0.1503$)) were not. Similar to the previous analysis, mitoses could not be evaluated. The prevalence of malignant characteristics in malignant and benign (stented and non-stented) bile duct brushings is summarized in Table 1.

Cytologic Findings Using Agreement by all 7 Reviewers as the Gold Standard

When agreement between all 7 reviewers was considered the gold standard, 4 malignant characteristics were statistically significantly associated with a malignant diagnosis (hypercellularity ($P=0.0384$), change in chromatin pattern ($P=0.0384$), nuclear irregularity ($P=0.0384$), pleomorphism ($P=0.0384$)) (Table 1). Nuclear molding or hugging, necrosis, and mitoses could not be evaluated because there were

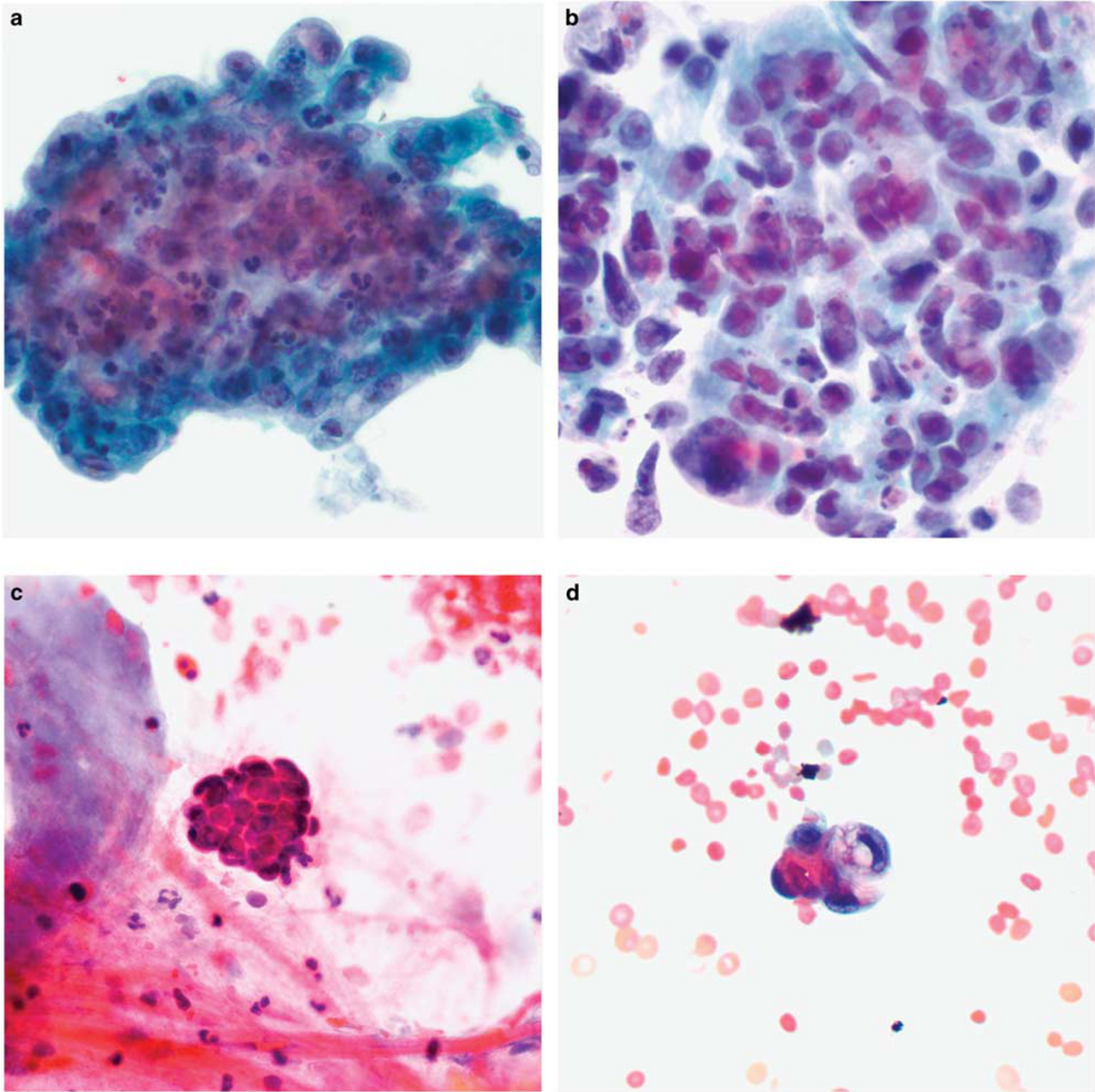


Figure 7 (a) Drunken honeycomb sheet of crowded malignant cells with hypochromasia, irregular nuclei, prominent nucleoli and infiltrating neutrophils (Papanicolaou stain, magnification $\times 200$). (b) This loose cluster of malignant cells shows marked nuclear irregularity, hyperchromasia, prominent nucleoli and single cell necrosis (Papanicolaou stain, magnification $\times 200$). (c,d) These 3-dimensional cell clusters show prominent nuclear molding/hugging (Papanicolaou stain, magnification $\times 200$ (c) and magnification $\times 400$ (d)).

no samples where all 7 reviewers agreed on the presence of these characteristics.

Odds of Malignancy Based on Presence of Each Cytologic Characteristic

Using agreement by 2/3 cytopathologists, odds ratios could not be calculated for 5 of the 14 characteristics because they were not present in both benign

and malignant cases. When odds ratios were calculated for the remaining 9 malignant characteristics, all were > 1.00 (range 1.30–47.46), indicating increased odds of malignancy if that specific characteristic was identified in a sample (Table 2). Nuclear pleomorphism had the highest odds ratio (OR): 47.46; 95% confidence interval (95% CI): 5.64, 399.29. The odds ratios for inflammation and necrosis were not statistically significant.

Table 1 Prevalence of malignant characteristics in benign and malignant bile duct brushings

| Characteristic | All malignant cases % (n = 30) | | | All benign cases % (n = 30) | | | Benign cases without stent % (n = 15) | | | P-value | |
|--|--------------------------------|-----|-----|-----------------------------|-----|-----|---------------------------------------|-----|-----|---------|--------|
| | Number of reviewers | | | Number of reviewers | | | Number of reviewers | | | | |
| | 2/3 CYTO | 4/7 | 7/7 | 2/3 CYTO | 4/7 | 7/7 | 2/3 CYTO | 4/7 | 7/7 | | |
| Hypochromasia/hyperchromasia (change in chromatin) | 70 | 13 | 13 | 10 | 7 | 0 | 13 | 7 | 7 | <0.0001 | 0.0384 |
| Nuclear irregularity | 67 | 70 | 13 | 13 | 10 | 0 | 20 | 7 | 7 | <0.0001 | 0.0384 |
| Nuclear pleomorphism | 62 | 57 | 13 | 3 | 62 | 0 | 7 | 0 | 0 | <0.0001 | 0.0384 |
| 3-D architecture | 52 | 63 | 3 | 3 | 3 | 0 | 7 | 0 | 0 | <0.0001 | 0.3132 |
| 2-Cell population | 57 | 60 | 7 | 0 | 3 | 0 | 0 | 0 | 0 | <0.0001 | 0.1503 |
| High nuclear to cytoplasmic ratio (> 50%) | 48 | 40 | 3 | 3 | 3 | 0 | 7 | 0 | 0 | <0.0001 | 0.3132 |
| Cytoplasmic mucin vacuoles | 43 | 43 | 7 | 13 | 3 | 0 | 13 | 13 | 0 | 0.0099 | 0.1503 |
| Cellular discohesion | 38 | 33 | 3 | 3 | 0 | 0 | 7 | 0 | 0 | 0.0010 | 0.3132 |
| Infiltrating inflammation | 43 | 37 | 7 | 33 | 27 | 10 | 40 | 33 | 27 | 0.4237 | 0.6404 |
| Hypercellularity | 23 | 30 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0048 | 0.0384 |
| Prominent nucleoli | 21 | 20 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0085 | 0.3132 |
| Nuclear molding or hugging | 20 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0099 | — |
| Necrosis | 17 | 13 | 0 | 13 | 0 | 0 | 13 | 0 | 13 | 0.7188 | — |
| Abnormal mitoses | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | — | — |

^aP-value for comparison of prevalence in malignant (n = 30) vs benign cases (n = 30) when 2/3 cytopathologists.

^b≥4/7 reviewers.

^c7/7 reviewers agreed on diagnosis.

Prevalence Score of Malignant Characteristics in Malignant and Benign Bile Duct Brushings

When each malignant characteristic that was present was given a numerical weight of ‘1’ point, based on agreement between 2/3 cytopathologists, the optimum number of characteristics needed to achieve an accurate diagnosis was investigated and was found to be ‘≥ 3’ (Table 3). In this analysis, characteristics that were not significantly associated with a malignant diagnosis (inflammation, necrosis and mitoses) were excluded so the score had a possible range of 0–11. Additionally, when various score cut points were evaluated, the ‘≤ 2 compared with ≥ 3’ malignant characteristics cut point resulted in the best combination of sensitivity (70%), specificity (97%), and accuracy (83%), but above that cut point (as the number of characteristics increased) overall specificity increased but accuracy and sensitivity decreased (Table 4). Of the 22 bile duct brushings with ≥ 3 malignant characteristics, only 1 patient’s sample turned out to be benign (5%) on long-term follow-up. This patient was a 70-year-old female in whom 2 of 3 cytopathologists and one GI pathologist identified 7 malignant characteristics (Figure 8). However, despite prolonged follow-up of 55 months she was alive with no clinical or radiologic evidence of pancreaticobiliary tract malignancy. In retrospect, we now conclude that this case represents one of the mimickers of carcinoma that we now refer to as detachment atypia. The features of this type of atypia elucidated by this study include clustered or crowded epithelial cells, without true 3-dimensionality and with increased nuclear to cytoplasmic ratio sometimes approaching, but not exceeding 50%. Additionally in these examples single intact cells with increased nuclear to cytoplasmic ratio cells, with smooth nuclear contours, hyperchromasia and small but distinct nucleoli are also seen (Figure 9).

Twenty-three of 30 (77%) benign samples had none of the 11 statistically significant malignant characteristics. Interestingly, 7 (23%) malignant bile duct brushings lacked all 11 characteristics when agreement by 2/3 cytopathologists was used as the gold standard (Table 3). When these 7 cases were later blindly re-reviewed by all 7 reviewers malignancy characteristics were indeed identifiable to a variable degree by several reviewers. Indeed, 3 or more malignant characteristics were identified by at least 2 reviewers in all 7 cases (range 3–8 characteristics/case). Re-review of the two cases that initially had only 1 malignant characteristic (per agreement between 2 of 3 cytopathologists) revealed 4–9 characteristics (mean of 6 characteristics) in one case and only 1 characteristic (mucin vacuoles) in the other. It is likely that although the latter case had a malignant biopsy there were no malignant cells on the bile duct brushing.

Using different gold standards (agreement by 2/3 cytopathologists vs agreement by ≥ 4/7 reviewers) the ‘≥ 3’ cut point remained the best one with

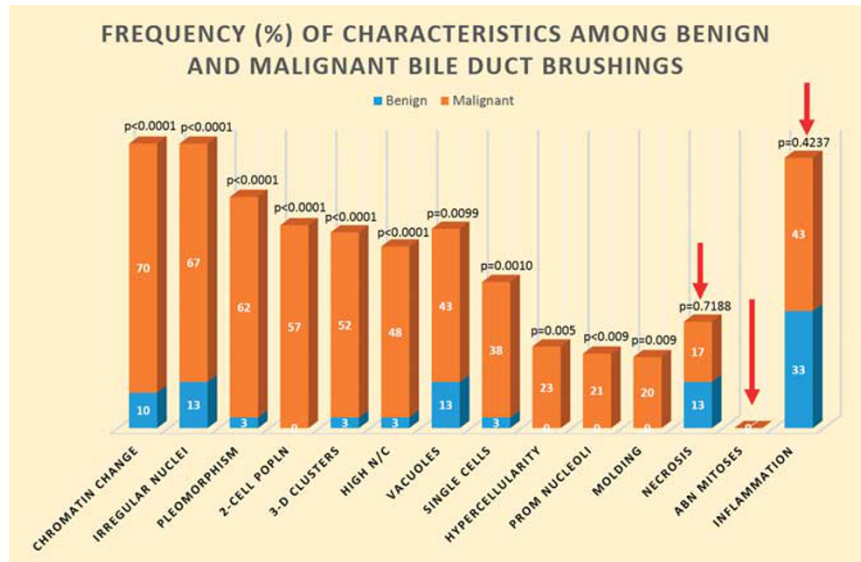


Figure 8 Using agreement by 2 of 3 cytopathologists as the initial gold standard 11 of 14 cytologic characteristics emerged as being statistically significantly more prevalent in malignant vs benign bile duct brushings, except for necrosis, abnormal mitoses and inflammation (red arrows) which showed no significant difference between benign and malignant cases.

Table 2 Most helpful characteristics based on greatest odds of malignancy

| Characteristic | Using agreement by 2/3 cytopathologists | | | Using agreement by 4/7 reviewers | | |
|---|---|-------------|----------|----------------------------------|-------------|----------|
| | Odds ratio | 95% CI | P-value | Odds ratio | 95% CI | P-value |
| 3-D architecture | 31.06 | 3.72–259.28 | 0.0015 | 50.09 | 5.97–420.36 | 0.0003 |
| Nuclear pleomorphism | 47.46 | 5.64–399.29 | 0.0004 | — | — | — |
| 2-Cell population ^a | — | — | — | 43.50 | 5.21–363.52 | 0.0005 |
| Chromatin pattern changes | 21.00 | 5.05–87.37 | < 0.0001 | 32.67 | 6.38–167.27 | < 0.0001 |
| High nuclear to cytoplasmic ratio (> 50%) | 27.06 | 3.24–225.92 | 0.0023 | 19.33 | 2.31–161.54 | 0.0063 |
| Cytoplasmic mucin vacuoles | 4.97 | 1.39–17.82 | 0.014 | 22.17 | 2.66–184.74 | 0.0042 |
| Irregular nuclear contours | 13.00 | 3.55–47.98 | 0.0001 | 21.00 | 5.05–87.37 | < 0.0001 |
| Cellular discohesion | 17.72 | 2.11–149.11 | 0.0082 | — | — | — |
| Less helpful characteristics^b | | | | | | |
| Hypercellularity ^a | — | — | — | — | — | — |
| Prominent nucleoli ^a | — | — | — | — | — | — |
| Nuclear molding or hugging ^a | — | — | — | — | — | — |

Abbreviation: BDB, bile duct brushing.

^aORs and CIs cannot be estimated when no sample in one or both of final outcome groups had the characteristic of interest.

^bThese characteristics were less helpful because they had incalculable odds ratios secondary to lower frequency in malignant brushings and/or complete absence in benign brushings.

identical sensitivity (70%), specificity (97%) and accuracy (83%) rates for both gold standards (Table 4). When agreement between 2/3 cytopathologists was used as the gold standard, the odds of malignancy increased 1.82-fold (95% CI: 1.29, 2.26) with each additional parameter. This figure was increased to 2.14-fold (95% CI: 1.36, 3.36) when agreement by 4 or more reviewers was analyzed as the gold standard. Based on the identified cut point of 3 characteristics, backwards elimination identified chromatin pattern, nuclear irregularity, and pleomorphism as the three characteristics that may be the most useful in determining a diagnosis (all three P-values < 0.001). Based on the gold standard of agreement between 4/7 reviewers, the sensitivity for the presence of all three characteristics was 57%,

the specificity was 100%, and the accuracy was 78%. When agreement between 2/3 cytopathologists was used as the gold standard, the sensitivity was 43%, the specificity was 97%, and the accuracy was 70%. These numbers proved to be lower than when any 3 of 11 parameter was utilized.

Reviewer Performance Using the Presence of ≥ 3 Malignant Characteristics

We also sought to determine if there was a difference in performance between experienced (n=4) and inexperienced (n=3) reviewers when utilizing malignant characteristics, based on the identification of 3 or more malignant characteristics cut point. The most

experienced reviewers were the board-certified cytopathologists (reviewers 1, 2 and 6) and the oncologic pathologist who routinely signed out pancreatobiliary cytology samples (reviewer 4, VoA). The least experienced reviewers were defined as the reviewers who rarely examined pancreatobiliary cytology specimens including one GI pathologist, the genitourinary pathologist and the cytopathology fellow with 3 months of training (reviewers 3, 5, and 7). Using a cut point of identification of ≥ 3 malignant characteristics, more experienced reviewers had a mean sensitivity of 69% (range 60–77%), specificity of 85% (range 77–93%) and accuracy of 77% (range 74–80%). Less experienced reviewers had a higher mean sensitivity of 73% (range 67–83%), lower mean specificity of 77% (range 66–90%) and slightly lower accuracy of 75% (range 73–78%) compared with more experienced reviewers. Individual reviewer performance is summarized in Table 5.

Discussion

While the identification of malignant characteristics in bile duct brushings would theoretically appear to

Table 3 Prevalence scores of malignant characteristics in malignant and benign bile duct brushings

| Score ^a | Malignant n (%) | Benign n (%) | P-value |
|--------------------|-----------------|---------------------|----------|
| 0 | 7 (23) | 23 (77) | < 0.0001 |
| 1 | 2 (33) | 4 (67) | 0.2501 |
| 2 | 0 (0) | 2 (100) | — |
| 3 | 2 (100) | 0 (0) | — |
| 4 | 2 (100) | 0 (0) | — |
| 5 | 1 (100) | 0 (0) | — |
| 6 | 2 (100) | 0 (0) | — |
| 7 | 2 (67) | 1 (33) ^b | 0.2501 |
| 8 | 8 (100) | 0 (0) | < 0.0001 |
| 9 | 3 (100) | 0 (0) | — |
| 10 | 1 (100) | 0 (0) | — |

^aScore calculation excludes inflammation, necrosis, and abnormal mitoses.

^bSee Figure 8 for illustrations of this case.

be a straightforward process, it is clearly a challenge for pathologists, and for general surgical pathologists in particular. While ancillary studies such as UroVysion FISH and immunohistochemistry are promising, they are not foolproof and are not widely available. Therefore, the identification of specific and reproducible morphologic characteristics should improve pathologist performance in identifying more malignant tumors.

We examined the strength of several well-established and some less well appreciated cytologic characteristics in identifying pancreatobiliary tract malignancy in bile duct brushings. Not surprisingly we found that, in order of (descending) proportion, change in chromatin pattern, nuclear irregularity, pleomorphism, 2-cell population, and 3-dimensional clusters were extremely helpful in accurately identifying malignancy in these specimens and were present in over 50% of cases, based on agreement between 2/3 cytopathologists. Other studies have also found nuclear irregularity and chromatin changes to be significantly helpful in identifying malignancy.^{13,16,19} High nuclear to cytoplasmic ratio and cellular discohesion (present in 38–48% of our cases) were also very helpful in our study. While prominent nucleoli are supportive of malignancy, they were not commonly identified in our malignant cases (20–28%). Layfield,¹ Nakajima¹⁵ and Waugh et al¹⁹ found that while prominent nucleoli were an important malignant characteristic on both conventional smears and ThinPrep, they were not always a very sensitive marker (20%), a finding that is similar to our own study. While malignant characteristics are important, if they are only rarely present in malignant samples then their helpfulness in accurate diagnosis becomes questionable. Hence, the prevalence of helpful ‘malignant’ characteristics is even more critical. This is exemplified by ‘abnormal mitoses’ which are a known malignant characteristic but were surprisingly only rarely seen (0–2%) in our malignant bile duct brushings. Another unexpected finding in our study was the similar frequency of

Table 4 Sensitivity, specificity, accuracy of malignant characteristics scores using various cut points

| Score ^a cut point | With agreement by 2/3 cytopathologists as gold standard | | | With agreement by ≥ 4 reviewers as gold standard | | |
|----------------------------------|---|-------------|----------|---|-------------|----------|
| | Sensitivity | Specificity | Accuracy | Sensitivity | Specificity | Accuracy |
| 0 compared with ≥ 1 | 77 | 77 | 77 | 80 | 87 | 83 |
| ≤ 1 compared with ≥ 2 | 70 | 90 | 80 | 70 | 93 | 82 |
| ≤ 2 compared with ≥ 3 | 70 | 97 | 83 | 70 | 97 | 83 |
| ≤ 3 compared with ≥ 4 | 63 | 97 | 80 | 63 | 97 | 80 |
| ≤ 4 compared with ≥ 5 | 57 | 97 | 77 | 60 | 97 | 78 |
| ≤ 5 compared with ≥ 6 | 53 | 97 | 75 | 53 | 100 | 77 |
| ≤ 6 compared with ≥ 7 | 47 | 97 | 72 | 40 | 100 | 70 |
| ≤ 7 compared with ≥ 8 | 40 | 100 | 70 | 33 | 100 | 67 |
| ≤ 8 compared with ≥ 9 | 13 | 100 | 57 | 17 | 100 | 58 |
| ≤ 9 compared with ≥ 10 | 3 | 100 | 52 | 13 | 100 | 57 |

^aScore calculation does not include inflammation, necrosis, and mitoses as these were not statistically significant.

what is interpreted by practitioners as 'necrosis' in benign and malignant cases, which suggests that it is not as specific a marker of malignancy as previously thought by some.⁷ Others have also reported similar results.^{13,14,16,19}

Inflammation also had a similar frequency in benign and malignant bile duct brushings, making it less useful in the accurate identification of malignant brushings. However, it should be noted that acute inflammation (a finding typically assumed to represent a benign, reactive inflammatory process) should not be ignored in all cases, particularly when associated with other concerning malignant characteristics. We have previously reported on the fact that pancreatic ductal adenocarcinomas, particularly its micropapillary and undifferentiated subtypes,

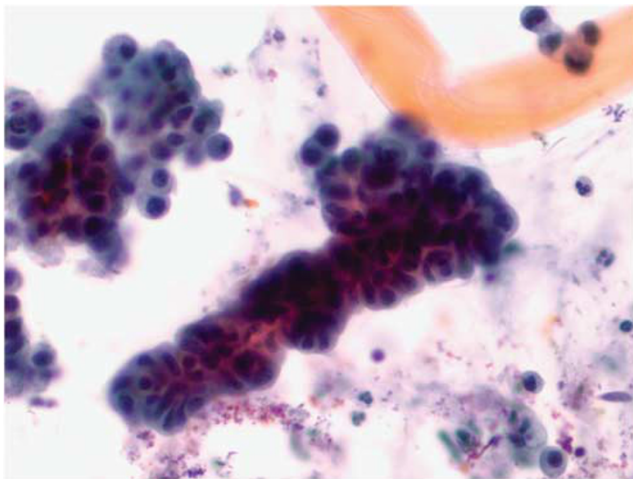


Figure 9 Benign bile duct brushing with detachment atypia. This example was called 'malignant' by 2/3 cytopathologists, based on their identification of 7 malignant characteristics including 3-dimensionality, discohesive single cells with high nuclear to cytoplasmic ratio and change in chromatin. Although these atypical cells appear clustered they lack true 3-dimensionality and have increased nuclear to cytoplasmic ratio, approaching, but not exceeding 50% (Papanicolaou stain, magnification $\times 200$). Rare single intact atypical cells with increased nuclear to cytoplasmic ratio, smooth nuclear contours, small but distinct nucleoli and hyperchromasia are present in the top right (Papanicolaou stain, magnification $\times 400$).

may demonstrate marked intra-epithelial infiltration by so-called 'tumor-infiltrating' neutrophils.^{25,26} Pancreatic ductal adenocarcinoma may involve the bile ducts resulting in 'positive' bile duct brushings. Thus, the presence of infiltrating neutrophils should not be ignored in such samples until other malignant characteristics have been clearly excluded. Among the benign bile duct brushings we also noted that non-stented benign brushings had a higher prevalence of inflammation than stented ones, suggesting that stent placement and associated subsequent management protocols suppress loco-regional inflammation, an expected treatment-related phenomenon.

The current study shows that although certain typically 'malignant' characteristics were seen in some of our benign bile duct brushings (albeit infrequently), this would suggest that some features, particularly when used single-handedly, are unreliable stand-alone criteria for distinguishing benign from malignant brushings, similar to some previously published observations.^{13,19} Nonetheless, we were able to define 11 cytomorphologic characteristics that are statistically significant in separating malignant from benign bile duct brushings among both experienced and inexperienced pathologists. In fact, for every 1-point increase in the number of malignant characteristics identified per case there was an incremental increase in the odds of malignancy, and brushings with 8 or more characteristics all turned out to be malignant on resection or follow-up. However, when so many characteristics are expected to be present in a sample then test sensitivity drops significantly with an increase in false-negatives. In fact, the optimum number of criteria that struck the critical balance in this study was 3. When 3 or more malignant characteristics were identified the sensitivity, specificity and accuracy were 70%, 97% and 83%, respectively. Additionally, the minimal gain in specificity with use of 4 or more characteristics was at significant cost to sensitivity and accuracy. Unfortunately, there was no ideal numerical combination of any specific cytologic criteria that brought additional strength to the diagnostic algorithm. This may not be surprising

Table 5 Reviewer performance using (≥ 3) score based on presence of malignant characteristics

| Reviewer | Overall performance | | | Concordance between (≥ 3) score and diagnosis |
|----------|---------------------|-----------------|--------------|--|
| | Sensitivity (%) | Specificity (%) | Accuracy (%) | ^a Concordance (%) |
| 1 | 67 | 93 | 80 | 98 |
| 2 | 70 | 77 | 74 | 84 |
| 3 | 83 | 66 | 74 | 93 |
| 4 | 77 | 80 | 78 | 90 |
| 5 | 70 | 76 | 73 | 95 |
| 6 | 60 | 90 | 75 | 95 |
| 7 | 67 | 90 | 78 | 93 |
| Range | 60–83 | 66–93 | 73–80 | 84–98 |
| Mean | 71 | 82 | 76 | 93 |

^aIf a reviewer identified ≥ 3 malignant characteristics, they were highly likely to diagnose the sample as malignant.

considering that adenocarcinomas manifest a variety of morphologic patterns in different patients. Additionally, when one considers the failed attempts at discovery of a 'magical numerical combination of criteria' for other extensively studied diseases including rheumatic fever, Kawasaki disease or papillary thyroid carcinoma, it is no surprise that bile duct brushings are not unique in this arena and failed to yield a specific combined set of criteria for optimal accuracy.

Tumor differentiation also has a critical role in the identification of malignancy. Poorly differentiated carcinoma is typically not a diagnostic challenge but well-differentiated tumors can be cytologically bland and extremely difficult to distinguish from benign and/or reactive changes morphologically, particularly in patients with primary sclerosing cholangitis, inflammation, prior stenting or even chronic pancreatitis, which is partly responsible for the test's low sensitivity. The Papanicolaou Society of Cytopathology recently published standardized reporting terminology for the sign out of pancreatobiliary cytology.²⁷ However, no specific cytologic criteria were defined for bile duct brushings in their study. Robins *et al* evaluated 19 cytologic criteria in pancreatic fine needle aspiration and suggested 3 major (overlapping nuclei, nuclear irregularity and chromatin clearing and/or clumping) and 4 minor (single epithelial cells, necrosis, mitoses and nuclear enlargement) criteria for improved sensitivity and specificity,²⁸ similar to the criteria our study elucidated. Their study was based on conventional smears, while most current bile duct brushings are made using liquid-based preparations. Waugh *et al* evaluated 100 ThinPrep bile duct brushings in search of useful malignant characteristics, and found that only nuclear features and patient age were statistically significant in separating benign from malignant cases.¹⁹ Additionally, 38% of their benign cases were misdiagnosed as malignant by at least one of four reviewers. Unlike Waugh *et al*,¹⁹ we found that cellular discohesion, 3-dimensional architecture and mucin vacuoles were additionally useful features and we included both ThinPrep and conventional smears in our study. We found that type of preparation had no significant effect on diagnosis.

While many studies used a small numbers of reviewers our study based the presence or absence of characteristics on agreement not only by three fellowship-trained cytopathologists but also agreement by non-cytopathologists who perform bile duct brushing evaluations in their daily practice, and found that the same criteria could be successfully applied by all seven pathologists, despite lack of formal cytopathology training. On the other hand, not surprisingly, experience played a role in reviewers' performance with more experienced reviewers having slightly higher specificity and accuracy rates than inexperienced ones. This implies that general surgical pathologists and others who less frequently sign out bile duct brushings could still be

trained to correctly identify characteristics, thus maximizing sensitivity, specificity and accuracy of diagnosis.

The search for improvements in diagnostic accuracy of bile duct brushings has led some to suggest triple testing (brush cytology, fluorescence *in situ* hybridization (FISH) and forceps biopsy) which has better accuracy, 82% sensitivity, 100% specificity, 100% positive predictive value, and 87% negative predictive value than brushing alone.^{29,30} However, these studies had sufficient specimen cellularity for additional testing, which is often not the case in clinical practice.^{1,31,32} UroVysion FISH has proven helpful in evaluating pancreatobiliary cytology specimens.^{14,29,33} Its' sensitivity (42.9%) was significantly higher than that of routine cytology (20.1%) ($P < 0.001$) and had identical specificity (99.6%).³⁴ A newer pancreatobiliary tract specific FISH probe that targets 1q21, 7p12, 8q24, and 9p21 detects significantly more cancers than UroVysion FISH (65% vs 46%; $P < 0.001$) with similar specificity (93% vs 91%).³⁵ However FISH poses technical and financial challenges. Immunohistochemistry is another useful adjuvant study in bile duct brushing assessment.³⁶ Over 50% of biliary cancers showed a maspin+/IMP3+/S100P+/pVHL- staining profile, and 20% showed a maspin+/IMP3-/S100P+/pVHL- profile in one study.³⁷ However immunohistochemistry lacks specificity, hence, identifying key morphologic characteristics as we did in this study will go a long way in improving test sensitivity without compromising specificity. Along those lines, in this cost-containment era, the assessment of 'risk of malignancy' by conventional cytology will likely be favored over other more expensive ancillary tests such as immunohistochemistry and FISH, hence the additional importance of more useful cytomorphologic criteria.

In conclusion, we evaluated specific cytologic criteria of malignancy in the diagnosis of carcinoma on bile duct brushings and found 11 malignant characteristics (hypo/hyperchromasia, nuclear irregularity, nuclear pleomorphism, presence of 2-cell population, 3-dimensional architecture, high nuclear cytoplasmic ratio (>50%), hypercellularity, cytoplasmic mucin, nuclear molding, discohesion, and presence of prominent nucleoli) to be very helpful in identifying malignancy. In fact, for every 1-point increase in number of characteristics, the odds of malignancy increased 2-fold. These characteristics were recognizable by both users with and without experience and training in cytopathology. Moreover, in evaluating the number of malignant characteristics, we determined that if ≥ 3 malignant characteristics were identified, sensitivity, specificity, and accuracy of a malignant diagnosis were maximized. However, we were unable to identify a set of three specific criteria that could be used in combination to improve diagnostic accuracy. The criteria we evaluated are equally applicable by both very experienced and less experienced pathologists. Hence

pathologists who infrequently see bile duct brushings can potentially be trained to correctly identify these characteristics so as to maximize the identification of malignancy in these samples.

Disclosure/conflict of interest

MDR received an honorarium from Medtronix for participation in a panel discussion. The remaining authors declare no conflict of interest.

References

- Logrono R, Kurtycz D, Molina C, *et al*. Analysis of false-negative diagnoses on endoscopic brush cytology of biliary and pancreatic duct strictures. *Arch Pathol Lab Med* 2000;124:387–392.
- Layfield LJ, Wax TD, Lee JG, *et al*. Accuracy and morphologic aspects of pancreatic and biliary duct brushings. *Acta Cytol* 1995;39:11–18.
- Burnett AS, Calvert TJ, Chokshi RJ. Sensitivity of endoscopic retrograde cholangiopancreatography standard cytology: 10-y review of the literature. *J Surg Res* 2013;184:304–311.
- 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;75:347–353.
- Adamsen S, Olsen M, Jendresen MB, *et al*. Endobiliary brush biopsy: Intra- and interobserver variation in cytological evaluation of brushings from bile duct strictures. *Scand J Gastroenterol* 2006;41:597–603.
- Duggan MA, Brasher P, Medlicott SA. ERCP-directed brush cytology prepared by the Thinprep method: test performance and morphology of 149 cases. *Cytopathology* 2004;15:80–86.
- 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;26:273–277.
- Hart J, Parab M, Mandich D, *et al*. IMP3 immunocytochemical staining increases sensitivity in the routine cytologic evaluation of biliary brush specimens. *Diagn Cytopathol* 2012;40:321–326.
- 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;53:1860–1865.
- 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;12:2793–2796.
- Papaxoinis K, Patsouris E, Athanassiadou P, *et al*. Contribution of nuclear morphometry by confocal laser scanning microscopy to the diagnosis of malignant bile duct strictures. *Acta Cytol* 2009;53:137–143.
- Stoos-Veic T, Bilic B, Kaic G, *et al*. Biliary brush cytology for the diagnosis of malignancy: a single center experience. *Coll Antropol* 2010;34:139–143.
- Cohen MB, Wittchow RJ, Johlin FC, *et al*. Brush cytology of the extrahepatic biliary tract: comparison of cytologic features of adenocarcinoma and benign biliary strictures. *Mod Pathol* 1995;8:498–502.
- 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;136:442–449.
- Nakajima T, Tajima Y, Sugano I, *et al*. Multivariate statistical analysis of bile cytology. *Acta Cytol* 1994;38:51–55.
- Jin YH, Kim SH, Park CK. Diagnostic criteria for malignancy in bile cytology and its usefulness. *J Korean Med Sci* 1999;14:643–647.
- Renshaw AA, Madge R, Jiroutek M, *et al*. Bile duct brushing cytology: statistical analysis of proposed diagnostic criteria. *Am J Clin Pathol* 1998;110:635–640.
- Henke AC, Jensen CS, Cohen MB. Cytologic diagnosis of adenocarcinoma in biliary and pancreatic duct brushings. *Adv Anat Pathol* 2002;9:301–308.
- Waugh MS, Guy CD, Maygarden SJ, *et al*. Use of the ThinPrep method in bile duct brushings: analysis of morphologic parameters associated with malignancy and determination of interobserver reliability. *Diagn Cytopathol* 2008;36:651–656.
- Salomao M, Gonda TA, Margolskee E, *et al*. Strategies for improving diagnostic accuracy of biliary strictures. *Cancer Cytopathol* 2015;123:244–252.
- Volmar KE, Vollmer RT, Routbort MJ, *et al*. Pancreatic and bile duct brushing cytology in 1000 cases: review of findings and comparison of preparation methods. *Cancer* 2006;108:231–238.
- 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;29:1358–1369.
- American Cancer Society Bile Duct Cancer (Cholangiocarcinoma). Available at <https://old.cancer.org/acs/groups/cid/documents/webcontent/003084-pdf.pdf> 2016. Accessed on 15 March 2017.
- Sullivan KM, Dean A, Soe MM. OpenEpi: a web-based epidemiologic and statistical calculator for public health. *Public Health Rep* 2009;124:471–474.
- Reid MD, Basturk O, Thirabanjasak D, *et al*. Tumor-infiltrating neutrophils in pancreatic neoplasia. *Mod Pathol* 2011;24:1612–1619.
- Khayyata S, Basturk O, Adsay NV. Invasive micropapillary carcinomas of the ampullo-pancreatobiliary region and their association with tumor-infiltrating neutrophils. *Mod Pathol* 2005;18:1504–1511.
- Pitman MB, Centeno BA, Ali SZ, *et al*. Standardized terminology and nomenclature for pancreatobiliary cytology: the Papanicolaou Society of Cytopathology guidelines. *Diagn Cytopathol* 2014;42:338–350.
- Robins DB, Katz RL, Evans DB, *et al*. Fine needle aspiration of the pancreas: In quest of accuracy. *Acta Cytologica* 1995;39:1–10.
- Nanda A, Brown JM, Berger SH, *et al*. Triple modality testing by endoscopic retrograde cholangiopancreatography for the diagnosis of cholangiocarcinoma. *Ther Adv Gastroenterol* 2015;8:56–65.
- Hartman DJ, Slivka A, Giusto DA, *et al*. Tissue yield and diagnostic efficacy of fluoroscopic and cholangioscopic techniques to assess indeterminate biliary strictures. *Clin Gastroenterol Hepatol* 2012;10:1042–1046.
- deBellis M, Sherman S, Fogel EL, *et al*. Tissue sampling at ERCP in suspected malignant biliary strictures. *Gastrointestinal endoscopy* 2002;56:11.

- 32 Ponchon T, Gagnon P, Berger F, *et al*. Value of endobiliary brush cytology and biopsies for the diagnosis of malignant bile duct stenosis: results of a prospective study. *Gastrointest Endosc* 1995;42: 565–572.
- 33 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;99:1675–1681.
- 34 Barr Fritcher EG, Kipp BR, Halling KC, *et al*. A multivariable model using advanced cytologic methods for the evaluation of indeterminate pancreatobiliary strictures. *Gastroenterology* 2009;136:2180–2186.
- 35 Barr Fritcher EG, Voss JS, Brankley SM, *et al*. An optimized set of fluorescence *in situ* hybridization probes for detection of pancreatobiliary tract cancer in cytology brush samples. *Gastroenterology* 2015;149: 1813–1824 e1811.
- 36 Fujisawa K, Maesawa C, Sato R, *et al*. Epigenetic status and aberrant expression of the maspin gene in human hepatobiliary tract carcinomas. *Lab Invest* 2005;85: 214–224.
- 37 Chen L, Huang K, Himmelfarb EA, *et al*. Diagnostic value of maspin in distinguishing adenocarcinoma from benign biliary epithelium on endoscopic bile duct biopsy. *Hum Pathol* 2015;46:1647–1654.