Integrated molecular pathology accurately determines the malignant potential of pancreatic cysts

Authors Mohammad A. Al-Haddad[1], Thomas Kowalski[2], Ali Siddiqui[2], Howard R. Mertz[3], Damien Mallat[4], Nadim Haddad[5], Nidhi Malhotra[5], Brett Sadowski[5], Mark J. Lybik[6], Sandeep N. Patel[7], Emuejevoke Okoh[7], Laura Rosenkranz[7], Michael Karasik[8], Michael Golioto[8], Jeffrey Linder[9], Marc F. Catalano[10]

Institutions Institutions are listed at the end of article.

submitted 26. March 2014
accepted after revision 28. August 2014

Bibliography
DOI http://dx.doi.org/10.1055/s-0034-1390742
Published online: 14.10.2014
Endoscopy 2015; 47: 136–142

ISSN 0013-726X

Corresponding author
Mohammad A. Al-Haddad, MD, MSc, FASGE, FACG
Division of Gastroenterology and Hepatology
Indiana University School of Medicine
550 N. University Blvd, Suite 4100
Indianapolis, IN 46202
USA
Fax: +1-317-948-8145
alhaddm@clevelandclinicabudhabi.ae

Background and study aims: Current diagnostic testing is inadequate to determine the malignant potential of pancreatic cysts, resulting in overcautious patient management. Integrated molecular pathology (IMP) testing combines molecular analysis with first-line test results (cytology, imaging, and fluid chemistry) to assess the malignant potential of pancreatic cysts. This multicenter study aimed to determine the diagnostic accuracy of IMP for pancreatic adenocarcinoma, and the utility of IMP testing under current guideline recommendations for managing pancreatic cysts.

Patients and methods: Patients who had undergone previous IMP testing as prescribed by their physician and for whom clinical outcomes were available from retrospective record review were included (n = 492). Performance was determined by correlation between clinical outcome and previous IMP diagnosis ("benign"/"statistically indolent" vs. "statistically higher risk [SHR]"/ "aggressive") or an International Consensus Guideline (Sendai 2012) criteria model for "surveillance" vs. "surgery." The Cox proportional hazards model determined hazard ratios for malignancy.

Results: Benign and statistically indolent IMP diagnoses had a 97% probability of benign follow-up for up to 7 years and 8 months from initial IMP testing. SHR and aggressive diagnoses had relative hazard ratios for malignancy of 30.8 and 76.3, respectively (both $P<0.0001$). Sendai surveillance criteria had a 97% probability of benign follow-up for up to 7 years and 8 months, but for surgical criteria the hazard ratio was only 9.0 ($P<0.0001$). In patients who met Sendai surgical criteria, benign and statistically indolent IMP diagnoses had a > 93% probability of benign follow-up, with relative hazard ratios for SHR and aggressive IMP diagnoses of 16.1 and 50.2, respectively (both $P<0.0001$).

Conclusion: IMP more accurately determined the malignant potential of pancreatic cysts than a Sendai 2012 guideline management criteria model. IMP may improve patient management by justifying more relaxed observation in patients meeting Sendai surveillance criteria. IMP can more accurately differentiate between the need for surveillance or surgery in patients meeting Sendai surgical criteria.

Introduction

Pancreatic cysts are being detected with increasing frequency due to advances in, and rising use of, imaging technology. A recent study suggests that only 2% of cysts are malignant at diagnosis [1]. The remaining patients have nonmalignant disease, with only a 0.4% chance of malignant transformation per year of surveillance [1]. Clinicians are thus faced with an epidemic of patients with pancreatic cysts who must be triaged to surveillance or surgery, the goal being to identify the minority of cysts with high malignant potential as early as possible while limiting unnecessary surgery.

Due to the high mortality rate of pancreatic cancer [2], International Consensus (Sendai), American College of Gastroenterology, and European Study Group on Cystic Tumours of the Pancreas treatment guidelines take a cautious approach and recommend resection for most cysts with any "worrisome" feature associated with malignancy [3–5]. However, the guideline-recommended criteria alone cannot accurately stratify patients for risk of malignancy, given that ~60%–80% of surgeries reveal nonmalignant disease [1,6–10]. Such a cautious management approach results in unnecessary morbidity [8]. Consequently, modalities that reduce aggressive treatment of indolent disease while facilitating early detection of cancer are needed to supplement current guidelines [11].

Molecular profiling of pancreatic cyst fluid has shown that high levels of high-quality DNA, loss

of heterozygosity of tumor suppressor genes, and *KRAS* point mutation correlate with malignancy [12,13]. However, none of these molecular features can individually determine malignancy risk. Similarly, the guideline-recommended tests of imaging, cytology, and fluid chemistry (amylase and/or carcinoembryonic antigen [CEA]) may identify features that guide decisions regarding surgery but cannot accurately determine the risk of malignancy individually, with the exception of a definitive malignant cytology result, which occurs in the minority of cases. In most cases, the integration of all molecular and clinical test results is required to provide an enhanced level of diagnostic and predictive information [13,14]. Integrated molecular pathology (IMP) testing of pancreatic cysts may be prescribed when patients lack definitive malignant cytology results, and combines molecular analysis with clinical test results to assess malignant potential.

The aim of the current study was to establish the clinical performance of IMP in diagnosing pancreatic adenocarcinoma in patients with pancreatic cysts that lack definitive malignant cytology results, and in risk-stratifying these patients for such malignancy. The clinical performance of a Sendai 2012 criteria [4] model for surgery or surveillance decisions, and the clinical utility of using IMP under these management guidelines, was also established in this study cohort. The ability of IMP testing to improve current management strategies was then assessed.

Patients and methods

Design

This study was an analysis of clinical outcomes data obtained from retrospective review of patient medical records documented in a National Pancreatic Cyst Registry, comprising 10 academic and private institutions in the United States. Approval was obtained at each site through the site-specific institutional review board or through Quorum IRB approval (#26022; Quorum Review IRB, Seattle, Washington, USA).

Patient population

All patients who had pancreatic cyst or associated duct fluid aspirate tested by IMP and were ≥ 18 years of age at the time of testing were eligible for inclusion. For patients who underwent follow-up, at least 23 months of follow-up imaging records were required unless clear benign or malignant clinical end points occurred within this timeframe (such patients were included). Exclusion criteria were: i) previous pancreatic cancer, ii) any treatment for pancreatic lesions prior to IMP testing, iii) presence of malignancy known prior to or during IMP diagnosis (to ensure blinding to actual patient outcome), iv) cases that could not be definitively categorized as having benign or malignant outcome, v) malignant events not related to the pancreas, and vi) diagnosis of neuroendocrine tumor by surgical pathology. Given that distinct areas of disease within the same patient can differ in malignant potential [15,16], cases in which the fine-needle aspiration (FNA) specimen tested was from a different area of the pancreas to where the diagnosis of malignancy was made were also excluded (e.g. malignant outcome diagnosed in a cystic lesion from the tail but FNA specimen tested was from a different cystic lesion located in the head).

IMP diagnosis

All IMP diagnoses were performed prior to the patients' inclusion in the study as a component of clinical testing according to the prescribing physician's standard of care, and were therefore all made blinded to the eventual clinical outcome. IMP testing was carried out by the same reference laboratory (PathFinderTG-Pancreas; RedPath Integrated Pathology, Inc., Pittsburgh, Pennsylvania, USA), and included a pathologist's interpretation of clinical features (imaging characteristics, levels of atypia and cellularity, amylase and/or CEA) integrated with results of molecular analysis of pancreatic cyst fluid or associated duct fluid aspirates collected by endoscopic ultrasound (EUS)-guided FNA, as per standard operating procedures (see supplementary material online). The algorithm for determining an IMP diagnosis is summarized in **Table 1**. Each case was categorized according to the four IMP diagnostic categories: "benign," "statistically indolent," "statistically higher risk (SHR)," and "aggressive."

As the diagnostic categories have evolved over time (see supplementary material online), all older cases were re-categorized according to current criteria using all imaging, atypia and cellularity, and amylase and/or CEA results available, to provide a consistent set of diagnoses indicative of the current standard performance for IMP. All re-categorizations were performed blinded to the final patient outcome.

Determination of clinical outcomes from patient records

Records from the time of initial IMP specimen collection onwards were reviewed, including imaging, endoscopy, surgery, surgical pathology, atypia (severe, moderate, mild, none, acellular), clinical laboratory results, IMP results, and oncology records if applicable. De-identified patient data were compiled in a secure online database (OpenClinica version 3.1.3; OpenClinica LLC, Waltham, Massachusetts, USA).

Clinical outcomes were categorized as "benign" or "malignant" (pancreatic adenocarcinoma). Benign outcomes included benign surgical pathology results, low-grade or intermediate-grade (including moderate) dysplasia, resolution of cyst (i.e. cyst not visible via imaging upon follow-up visit), or clinical follow-up by imaging (EUS, magnetic resonance imaging, computed tomography, endoscopic retrograde cholangiopancreatography) for ≥ 23 months without evidence of malignant outcome. This fol-

Table 1 Criteria for integrated molecular pathology diagnostic categories.

Diagnostic category	Molecular criteria[1]	Co-existing concerning clinical features[2]
Benign	DNA lacks molecular criteria	Not considered for this diagnosis
Statistically indolent	DNA meets 1 molecular criterion	None
SHR	DNA meets 1 molecular criterion	1 or more
Aggressive	DNA meets at least 2 molecular criteria	Not considered for this diagnosis

SHR, statistically higher risk.

[1] Four molecular criteria that have been independently correlated with pancreatic malignancy or high-grade dysplasia are used to make an integrated molecular pathology diagnosis: i) a single high-clonality mutation, ii) elevated level of high-quality DNA, iii) multiple low-clonality mutations; iv) a single low-clonality oncogene mutation [13].

[2] Include any of the following: cyst size > 3 cm, growth rate > 3 mm/year, duct dilation > 1 cm, carcinoembryonic antigen level > 1000 ng/mL, cytologic evidence of high-grade dysplasia.

low-up period was selected as a reasonable time for latent pancreatic adenocarcinoma to become evident. Malignant outcomes were determined by conclusive indications of malignancy, such as surgical pathology diagnosis of high-grade dysplasia, carcinoma in situ, or adenocarcinoma, malignant cytology results (unknown during IMP diagnosis; patients were excluded if known), clinically confirmed pancreatic cancer in patient records, and death attributed to pancreatic cancer. The date of malignant outcome was defined as the earliest date at which a definitive diagnosis of malignancy could be made.

Performance characteristics

The performance characteristics of IMP were evaluated by determining the correlation between blinded IMP diagnoses and actual clinical outcomes from medical record review. Accuracy, sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) for malignant outcome were calculated using 2×2 tables. The benign and statistically indolent categories were considered to be low malignant potential categories, and the SHR and aggressive categories were considered to be high malignant potential categories.

The performance of a model of the Sendai 2012 criteria for managing this same cohort of patients according to imaging, cytologic, and amylase and/or CEA results was determined in the same way. In the model, patients were categorized as having met Sendai 2012 "surveillance" criteria (low malignant potential) or Sendai 2012 "surgery" criteria (high malignant potential). The surgery category required at least one of the following features: presence of obstructive jaundice in a patient with a cystic lesion of the head, cyst size >3 cm, enhancing solid component or definite mural nodule confirmed by EUS, severe (i.e. suspicious) cytology, main duct involvement, main duct dilation of ≥1 cm, abrupt changes in duct caliber, or a presumptive diagnosis of mucinous cystic neoplasm based on cytology report indications of mucin. The surveillance category lacked all such features. This model does not match the published Sendai 2012 guidance exactly, as the patient cohort included in the study comprised only patients for whom IMP was ordered as part of standard clinical care (i.e. those with negative, nondiagnostic, indeterminate or acellular cytology results) and not all possible pancreatic cyst patients.

Statistical analysis

Statistical analysis was performed using R statistical software (r-project.org). Exact binomial tests were used to determine statistical significance and 95% confidence intervals (CI) for accuracy, sensitivity, specificity, NPV, and PPV. Comparisons between surgical outcome and all clinical outcome populations were performed using Fisher's exact test. Comparisons of sensitivity and specificity between IMP and the Sendai 2012 model were performed using McNemar's test, and comparisons of NPV and PPV between IMP and the Sendai model were made using a weighted generalized scoring statistic [17].

Malignancy-free outcome from the date of initial IMP diagnosis was estimated using Kaplan–Meier curves. The hazard ratio for risk of malignant outcome was calculated using the univariate Cox proportional hazards model for the SHR and aggressive IMP categories vs. the benign and statistically indolent categories combined. The same univariate Cox proportional hazards model was used to determine the hazard ratio between the surgery and surveillance categories of the Sendai model. For subcategory comparisons, such as aggressive vs. SHR, multiple comparison adjustment with Tukey's test was used to determine the hazard ratio. A post hoc, multivariate Cox proportional hazards analysis was also performed, adjusting for age, sex, symptoms related to pancreatic disease, patient history of cancer (nonpancreatic), and patient family history of any malignancy, to confirm the findings of the univariate analysis.

Results

Patients

The study was carried out between April 2011 and August 2013. Records for 1864 patients were reviewed (patients undergoing EUS-FNA and IMP testing between January 2005 and April 2013). Of these, 1372/1864 cases did not meet study inclusion criteria: 1280 had insufficient follow-up to confirm benign outcome (<23 months) and 92 met exclusion criteria. Overall, 492/1864 patients were included (338 women and 154 men; mean age at IMP testing 64.9 years). Of these, 468/492 IMP diagnoses were re-categorized prior to analysis, and 24/492 were already classified according to the current four IMP diagnostic categories. The sources of clinical outcomes with follow-up time are shown in Supplementary **Table e2** (available online). For outcomes based on imaging surveillance, follow-up of ≥3 years was available for 46% of patients. Pancreatic cyst clinical characteristics and surgical pathology classifications are shown in Supplementary **Table e3** and **Table e4** (available online).

Clinical performance characteristics of IMP

The overall accuracy, NPV, and specificity of IMP for malignant outcome were high (90%, 97%, and 91%, respectively) (**Table 5**). The sensitivity of IMP for malignant outcome was 83% and the PPV was 58% (**Table 5**). Overall accuracy of IMP in the subset of surgical outcomes was 10% lower than that in the all-outcomes population (Supplementary **Table e6**, available online). For Kaplan–Meier analysis, the median follow-up time for patients with benign and statistically indolent diagnoses was 35

Table 5 Performance characteristics of integrated molecular pathology diagnosis and the model of the international consensus guideline (Sendai 2012) criteria regarding ability to differentiate between cysts with and without malignant potential (n = 492).

	IMP % [95%CI]	Sendai 2012 model % [95%CI]	*P* for IMP vs. Sendai 2012 model
Accuracy	89.6 [86.6–92.2]	52.2 [47.7–56.7]	N/A
Sensitivity	83.3 [72.1–91.4]	90.9 [81.2–96.6]	0.17
Specificity	90.6 [87.4–93.2]	46.2 [41.4–51.1]	<0.0001
NPV	97.2 [95.1–98.6]	97.0 [93.7–98.9]	0.88
PPV	57.9 [47.3–68.0]	20.8 [16.2–25.9]	<0.0001
Positive likelihood ratio	8.9 [6.5–12.2]	1.7 [1.5–1.9]	<0.0001
Negative likelihood ratio	0.2 [0.1–0.3]	0.2 [0.1–0.4]	0.88

IMP, integrated molecular pathology; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; N/A, not applicable.

months (range 23 months – 7 years 8 months). Patients with these diagnoses had a 97% probability of a benign outcome for up to 7 years and 8 months from initial IMP testing (**Fig. 1**). There were only a few instances of malignant outcome (11/397) in the benign and statistically indolent categories, and all of these malignancies occurred within 23 months of initial IMP diagnosis (**Fig. 1**).

Malignant outcome was confirmed in most patients with aggressive IMP diagnoses (88%) and in 47% with SHR diagnoses (Supplementary **Table e6**, available online). Of the SHR and aggressive cases with malignant outcome, most were confirmed within 3 months of IMP testing (**Fig. 1**). The probability of a benign outcome <1 year after initial IMP testing was 35% in patients with SHR diagnoses and 9% in those with aggressive diagnoses, with univariate hazard ratios for malignant outcome of 30.8 and 76.3, respectively, relative to those with benign and statistically indolent diagnoses (both $P<0.0001$) (Supplementary **Table e7**, available online). The hazard ratio for aggressive diagnosis relative to SHR diagnosis was 2.4 ($P<0.01$, confirmed by Tukey's multiple comparison test). After adjusting for covariates, hazard ratios for the SHR (27.8) and aggressive (79.2) categories were also statistically significant (both $P<0.0001$) (Supplementary **Table e7**, available online).

Clinical performance characteristics of Sendai 2012 model

The performance characteristics of the Sendai 2012 model in this patient cohort are shown in **Table 5**. Overall accuracy of the Sendai model was 13% lower for the subset of surgical pathology outcomes only compared with all outcomes (Supplementary **Table e8**, available online). In Kaplan–Meier analysis, patients who met Sendai surveillance criteria had a 97% chance of having a benign outcome for up to 7 years and 8 months from initial IMP testing (**Fig. 2**). However, patients who met Sendai criteria for surgery also had a high probability (75%) of having a benign outcome at this time (**Fig. 2**). The univariate hazard ratio for malignant outcome was 9.0 for those who met Sendai surgical criteria relative to those who met surveillance criteria ($P<0.0001$) (Supplementary **Table e7**, available online). After adjusting for covariates, the hazard ratio for those who met surgical criteria remained statistically significant (8.1, $P<0.0001$) (Supplementary **Table e7**, available online).

Clinical utility of IMP testing in patients who met Sendai 2012 model criteria for surgery or surveillance

Comparison of IMP with the Sendai 2012 model showed similar sensitivity, NPV, and negative likelihood ratio ($P>0.17$) but statistically better specificity, PPV, and positive likelihood ratio for IMP (all $P<0.0001$). The majority of patients who met Sendai criteria for surgery actually had benign outcomes (79% [229/289]; Supplementary **Table e8**, available online). Of these patients meeting Sendai surgical criteria but having benign clinical outcomes, 84% (193/229) had an IMP diagnosis of benign or statistically indolent (Supplementary **Table e9**, available online).

Kaplan–Meier curves for the ability of IMP to further improve risk stratification of patients who met Sendai criteria for surgery are shown in **Fig. 3**. The probability of benign outcome at follow-up for benign and statistically indolent IMP diagnoses in patients who met Sendai criteria for surgery was 93% and 97%, respectively, even up to ~5 years after initial IMP testing. There was no statistical difference in the univariate hazard ratios between benign and statistically indolent diagnoses ($P>0.46$, confirmed by Tukey's multiple comparison test). By contrast, the probability of benign outcome at follow-up in SHR and aggressive IMP diagnoses was much lower (36% and 5%, respectively). In patients who met Sendai surgical criteria, the univariate hazard ra-

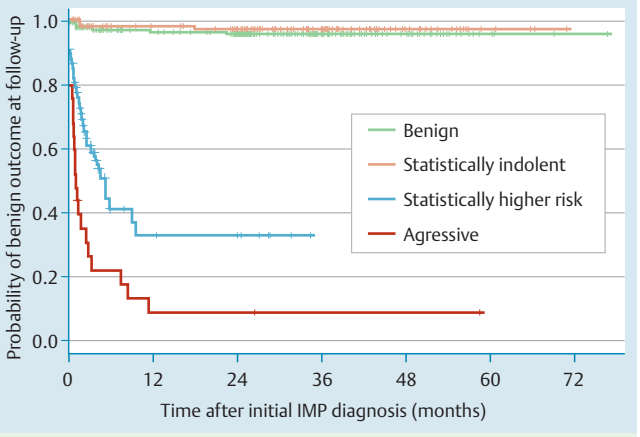

Fig. 1 Integrated molecular pathology (IMP) risk stratification for malignant potential in all patients (n = 492). Kaplan–Meier analysis for probability of a benign outcome at follow-up for each IMP diagnostic category.

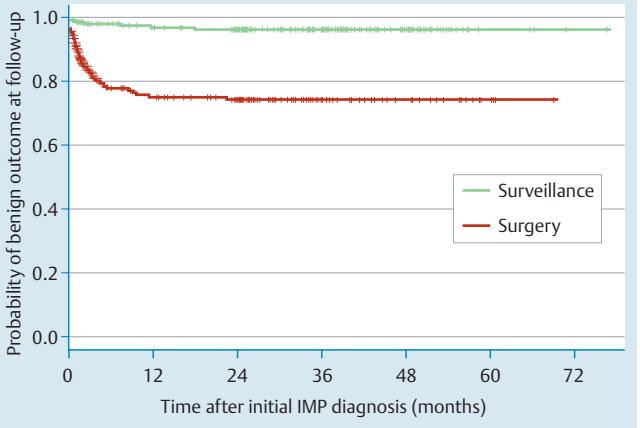

Fig. 2 Sendai 2012 risk stratification for malignant potential in all patients (n = 492). Kaplan–Meier analysis for probability of a benign outcome at follow-up in patients who met Sendai 2012 model criteria categories for surveillance or surgery.

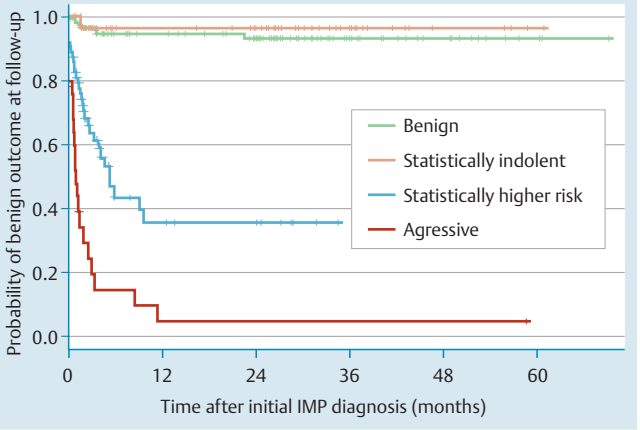

Fig. 3 Integrated molecular pathology (IMP) risk stratification for malignant potential in patients who met Sendai 2012 model criteria for surgery. Kaplan–Meier analysis for probability of benign outcome at follow-up for each IMP diagnostic category.

tios for malignant outcome were 16.1 (95 %CI 7.6 – 34.3) for SHR and 50.2 (95 %CI 22.6 – 111.5) for aggressive IMP diagnoses relative to benign and statistically indolent combined (both $P<0.0001$).

Of the patients who met Sendai surveillance criteria, 197 /203 had benign outcomes and 6/203 had malignant outcomes (Supplementary **Table e9**, available online). Surveillance criteria were confirmed by benign or statistically indolent IMP diagnoses in 98 % (193/197) of these patients. Of the six false-negative cases identified by Sendai criteria for surveillance, four were correctly diagnosed as SHR or aggressive disease by IMP.

Discussion

The greatest challenge in managing pancreatic cysts is the accurate stratification of patients at risk of malignancy so that unnecessary surgery is minimized. As none of the currently available guideline-recommended tests can accurately determine malignant potential in pancreatic cysts, additional strategies that provide reliable information regarding the presence, absence, or increased risk of malignancy are needed to help guide patient management.

This study showed that IMP provides clinically valid and useful diagnostic information that can improve management of patients with pancreatic cysts. The high NPV, specificity, and probability of follow-up with benign outcome indicate that benign and statistically indolent IMP diagnoses reliably predict benign disease. IMP therefore identifies patients in whom surgery is avoidable and surveillance is justified. Furthermore, in patients with benign or statistically indolent diagnoses, no cases of malignant outcome occurred beyond 2 years of follow-up from initial IMP diagnosis, which should help guide surveillance intervals. The sensitivity and PPV of IMP for malignant outcome indicates that aggressive and SHR IMP diagnoses are reliable predictors of malignancy. Moreover, the hazard ratios for risk of malignant outcome were significantly higher for the SHR and aggressive IMP categories relative to benign/statistically indolent, and for the aggressive category relative to SHR. IMP therefore presents a useful risk stratification tool for clinicians to differentiate benign and indolent cysts from those at higher risk of malignancy that require close surveillance or immediate surgery.

The application of the Sendai 2012 model to this patient cohort emphasizes a well-documented problem: too many patients undergo surgery for benign pancreatic cysts [1,6 – 10]. The NPV, negative likelihood ratio, and sensitivity of the Sendai criteria were high, as was the probability of benign outcome at follow-up. However, the accuracy, specificity, positive likelihood ratio, and PPV were low due to the high number of false-positive results for malignant outcome when applying surgical criteria. Consequently, patients meeting Sendai criteria for surgery had only a 9.0-times higher risk of malignant outcome, which would result in many surgeries on benign or indolent cysts. Nearly identical results were observed when European treatment guideline criteria for surgery and surveillance [5] were evaluated (data not shown). Of note, performance of the Sendai 2012 criteria was examined only in patients for whom IMP testing had been prescribed (those with negative, nondiagnostic, indeterminate or acellular cytology results), rather than in all possible pancreatic cyst patients. Thus, the study cohort did not include those with the “high-risk stigmata” of malignant cytology prior to IMP testing, which is an absolute indication for surgery according to current guidance, because additional information is not needed in these patients to determine whether surgery is required.

Although the Sendai 2012 guidance refers specifically to the management of intraductal papillary mucinous neoplasm and mucinous cystic neoplasm, we believe that from the perspective of a practicing physician operating in the clinic, all cysts are considered mucinous until proven otherwise. Given that the Sendai 2012 guidance is intended to manage surveillance or surgery decisions, we think it is reasonable to assume that patients who do not have mucinous cystic neoplasm or other intraductal papillary mucinous neoplasm criteria for surgery would undergo surveillance, which is how the Sendai 2012 criteria were modeled in the current analysis, and this is reflected in the European guidelines. Also, we considered a cyst size of > 3 cm to be an indication for surgery, but recognize that the Sendai 2012 guidance is slightly more complex with respect to cyst size. To address this difference, the analysis was repeated with complete exclusion of cyst size from the surveillance and surgery categories; this had minimal impact on the performance of the Sendai 2012 model in this cohort (only 8.0-times higher risk of malignant outcome for patients meeting surgical criteria; data not shown). Furthermore, in this cohort, the addition of CEA ≥ 192 ng/mL as a criterion for mucin, and thus surgery, decreased the accuracy of the Sendai model due to an increase in false-positive cases without any reduction in false-negative cases (only 5.9-times higher risk of malignant outcome for patients meeting surgical criteria; data not shown). Omitting all presumptive cytologic and CEA criteria for mucinous cysts from the Sendai model only slightly improved risk stratification due to a modest reduction in false-positive cases (11.6-times higher risk of malignant outcome for patients meeting surgical criteria; data not shown). Regardless of these modeling limitations, the performance of the Sendai criteria in our model paralleled that of another study with a larger cohort (n = 767) and similar follow-up time (PPV 23 % and NPV 99.5 %) [18].

Comparison of the performance characteristics of IMP with those of the Sendai 2012 model indicated similar sensitivity, negative likelihood ratio, and NPV for malignancy and similar probability of benign outcome at follow-up. By contrast, there were statistically significant differences in specificity, positive likelihood ratio, and PPV favoring IMP due to a reduction in false-positive cases. Such differences were further emphasized when Sendai criteria and IMP risk stratification capabilities were compared. Patients who had high malignant potential per IMP diagnoses had a much higher risk of malignant outcome than those who met Sendai criteria for high malignant potential.

To examine the clinical utility of using IMP under current guideline-recommended management strategies, IMP performance was evaluated in patients meeting Sendai 2012 criteria for surgery and for surveillance. In patients meeting Sendai surgical criteria, benign and statistically indolent IMP diagnoses correctly predicted benign outcomes in the majority, with a > 93 % probability of benign outcome at follow-up. Comparatively, patients with SHR and aggressive diagnoses were at higher risk of malignant outcome. Thus, IMP can more accurately stratify the malignant potential than guideline criteria for surgery. Although the majority of patients meeting Sendai 2012 criteria for surveillance had benign outcomes, some false-negatives were present. IMP was not only able to confirm surveillance in nearly all of these patients but was also able to correctly identify most false-negative cases. IMP therefore improves guideline-recommended management strategies by justifying more relaxed observation in pa-

tients likely to have benign disease course, and closer surveillance or surgery in patients at higher risk of pancreatic adenocarcinoma.

The primary limitation of this study is the retrospective nature of the outcomes data, which are subject to the drawbacks inherent to this type of study, including insufficient or inaccessible documentation, which resulted in many cases not meeting the prespecified inclusion criterion of follow-up ≥ 23 months. Data that would be used for the categorization of patients according to Sendai 2012 criteria were also not specified in a significant proportion of patients, as the collection of information was started prior to publication of the 2012 guidelines. The cohort examined also precluded assessment of the Sendai surgical criterion of malignant cytology, because patients with initial malignant cytology results do not require IMP testing. However, as noted above, an independent study showed very similar results to those reported here when malignant cytology was included [18]. Finally, considering the nature of these lesions, the mean follow-up period for benign disease in this study is too short for firm conclusions to be made beyond 3 years; similar analyses with longer follow-up are desirable.

Retrospective record review is currently the only feasible method for assessing more recent diagnostic criteria or tests due to the benign nature of most pancreatic cysts. The reference standard in such a study is surgical pathology. However, surgery-only populations do not reflect the total patient population for which diagnostic criteria are intended, which includes patients who will have a benign disease course and thus, in reality, undergo surveillance. An advantage of the current study is that it examined performance of diagnostic criteria in a population comprising both surgery and surveillance subpopulations, thus providing a more real-life evaluation for the performance of both guideline criteria and IMP testing. The overall accuracy of both IMP and the Sendai 2012 model was statistically lower for the analysis of surgical outcomes only compared with the analysis of all types of outcome.

In conclusion, an IMP diagnosis of benign or statistically indolent is a highly reliable predictor that a pancreatic cyst is benign and has low potential for malignant transformation over a median follow-up of ~3 years in patients who lack definitive malignant cytology results. Furthermore, SHR and aggressive IMP diagnoses identify patients who are at significantly higher risk of pancreatic adenocarcinoma. IMP is more accurate than Sendai 2012 criteria in risk stratifying patients according to their potential for pancreatic adenocarcinoma. IMP may therefore improve guideline-recommended patient management strategies by increasing confidence that observation is more appropriate in the majority of patients, who are likely to have a benign disease course. In a minority of patients, closer surveillance or surgical consultation may be necessary due to significantly higher risk of pancreatic adenocarcinoma. IMP is therefore able to limit the overtreatment of inconsequential disease while accurately detecting cancer, which is an essential characteristic of a clinically useful diagnostic test [11].

Competing interests: M. A.-H. received a grant for this research from RedPath to his institution during the conduct of the study, and received personal fees from AbbVie, Boston Scientific, and Forest outside of the submitted work. M.K. received a research grant for this research to his institution from RedPath during the conduct of the study, and acknowledges support to his institution from The Ron Foley Foundation (www.ronsrun.org), West Hartford, Connecticut, outside of the submitted work. M.F.C. received a grant for this research from RedPath to his institution during the conduct of the study. The remaining authors have no conflicts to disclose.

Institutions
[1] Department of Medicine, Indiana University, Indianapolis, Indiana, USA
[2] Department of Medicine, Jefferson Digestive Disease Institute, Thomas Jefferson University, Philadelphia, Pennsylvania, USA
[3] Nashville GI Associates, Nashville, Tennessee, USA
[4] Premier Gastroenterology of Texas, Dallas, Texas, USA
[5] Division of Gastroenterology, MedStar Georgetown University Hospital, Washington, DC, USA
[6] Northside Gastroenterology, Indianapolis, Indiana, USA
[7] Department of Medicine, University of Texas San Antonio, San Antonio, Texas, USA
[8] Connecticut GI, Hartford, Connecticut, USA
[9] Digestive Health Associates of Texas, Dallas, Texas, USA
[10] Wisconsin Center for Advanced Research, St. Luke's Medical Center, Milwaukee, Wisconsin, USA

Acknowledgments

Mohammad Al-Haddad would like to acknowledge Jennifer Schaffter Stuart for her contributions to this study. Michael Karasik would like to acknowledge Leah Ehle, Advanced Practice Registered Nurse, for her contributions to the study. Statistical analyses by Yi Zhang, PhD, and writing assistance from Helen Varley, PhD, CMPP (Envision Custom Solutions, Horsham, UK), were supported by RedPath.

References

1 *Wu BU*, *Sampath K*, *Berberian CE* et al. Prediction of malignancy in cystic neoplasms of the pancreas: a population-based cohort study. Am J Gastroenterol 2014; 109: 121 – 129
2 *Siegel R*, *Naishadham D*, *Jemal A*. Cancer statistics, 2013. CA Cancer J Clin 2013; 63: 11 – 30
3 *Khalid A*, *Brugge W*. ACG practice guidelines for the diagnosis and management of neoplastic pancreatic cysts. Am J Gastroenterol 2007; 102: 2339 – 2349
4 *Tanaka M*, *Fernandez-del Castillo C*, *Adsay V* et al. International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas. Pancreatology 2012; 12: 183 – 197
5 *Del Chiaro M*, *Verbeke C*, *Salvia R* et al. European experts consensus statement on cystic tumours of the pancreas. Dig Liver Dis 2013; 45: 703 – 711
6 *Grobmyer SR*, *Cance WG*, *Copeland EM* et al. Is there an indication for initial conservative management of pancreatic cystic lesions? J Surg Oncol 2009; 100: 372 – 374
7 *Allen PJ*, *D'Angelica M*, *Gonen M* et al. A selective approach to the resection of cystic lesions of the pancreas: results from 539 consecutive patients. Ann Surg 2006; 244: 572 – 582
8 *Gaujoux S*, *Brennan MF*, *Gonen M* et al. Cystic lesions of the pancreas: changes in the presentation and management of 1,424 patients at a single institution over a 15-year time period. J Am Coll Surg 2011; 212: 590 – 600
9 *Spinelli KS*, *Fromwiller TE*, *Daniel RA* et al. Cystic pancreatic neoplasms: observe or operate. Ann Surg 2004; 239: 651 – 657
10 *Huang ES*, *Turner BG*, *Fernandez-Del-Castillo C* et al. Pancreatic cystic lesions: clinical predictors of malignancy in patients undergoing surgery. Aliment Pharmacol Ther 2010; 31: 285 – 294
11 *Esserman LJ*, *Thompson IMJr*, *Reid B*. Overdiagnosis and overtreatment in cancer: an opportunity for improvement. JAMA 2013; 310: 797 – 798
12 *Khalid A*, *McGrath KM*, *Zahid M* et al. The role of pancreatic cyst fluid molecular analysis in predicting cyst pathology. Clin Gastroenterol Hepatol 2005; 3: 967 – 973
13 *Khalid A*, *Zahid M*, *Finkelstein SD* et al. Pancreatic cyst fluid DNA analysis in evaluating pancreatic cysts: a report of the PANDA study. Gastrointest Endosc 2009; 69: 1095 – 1102
14 *Shen J*, *Brugge WR*, *Dimaio CJ* et al. Molecular analysis of pancreatic cyst fluid: a comparative analysis with current practice of diagnosis. Cancer 2009; 117: 217 – 227

15 *Matthaei H*, *Norris AL*, *Tsiatis AC* et al. Clinicopathological characteristics and molecular analyses of multifocal intraductal papillary mucinous neoplasms of the pancreas. Ann Surg 2012; 255: 326 – 333
16 *Lee YH*, *Lee NJ*, *Kim JH* et al. Sonographically guided fine needle aspiration of thyroid nodule: discrepancies between cytologic and histopathologic findings. J Clin Ultrasound 2008; 36: 6 – 11
17 *Kosinski AS*. A weighted generalized score statistic for comparison of predictive values of diagnostic tests. Stat Med 2013; 32: 964 – 977
18 *Lawson R*, *Savides T*, *Kwong W* et al. Natural history of pancreatic cystic lesions: validation of the 2012 Sendai criteria. Am J Gastroenterol 2013; 108: S91

Supplementary methods, Table e2, e3, e4 and e6, e7, e8, e9

online content viewable at: www.thieme-connect.de