Diagnosing Chronic Pancreatitis

Comparison and Evaluation of Different Diagnostic Tools

Yama Issa, MD, PhD, Hjalmar C. van Santvoort, MD, PhD,† Susan van Dieren, PhD,* Marc G. Besselink, MD, PhD,* Marja A. Boermeester, MD, PhD,* and Usama Ahmed Ali, MD*‡*

Objectives: This study aims to compare the M-ANNHEIM, Büchler, and Lüneburg diagnostic tools for chronic pancreatitis (CP).

Methods: A cross-sectional analysis of the development of CP was performed in a prospectively collected multicenter cohort including 669 patients after a first episode of acute pancreatitis. We compared the individual components of the M-ANNHEIM, Büchler, and Lüneburg tools, the agreement between tools, and estimated diagnostic accuracy using Bayesian latent-class analysis.

Results: A total of 669 patients with acute pancreatitis followed-up for a median period of 57 (interquartile range, 42–70) months were included. Chronic pancreatitis was diagnosed in 50 patients (7%), 59 patients (9%), and 61 patients (9%) by the M-ANNHEIM, Lüneburg, and Büchler tools, respectively. The overall agreement between these tools was substantial ($\kappa = 0.75$). Differences between the tools regarding the following criteria led to significant changes in the total number of diagnoses of CP: abdominal pain, recurrent pancreatitis, moderate to marked ductal lesions, endocrine and exocrine insufficiency, pancreatic calcifications, and pancreatic pseudocysts. The Büchler tool had the highest sensitivity (94%), followed by the M-ANNHEIM (87%), and finally the Lüneburg tool (81%).

Conclusions: Differences between diagnostic tools for CP are mainly attributed to presence of clinical symptoms, endocrine insufficiency, and certain morphological complications.

Key Words: chronic pancreatitis, diagnostic tools, diagnosis, M-ANNHEIM, Büchler, Lüneburg

(*Pancreas* 2017;46: 1158–1164)

An accurate diagnostic tool is important for the diagnosis, treatment, and follow-up of patients with (suspected) chronic pancreatitis (CP). This could lower the burden of additional diagnostic examinations and unnecessary resource utilization and allow for timely treatment.[1,2] There is, however, much controversy about the diagnosis of CP. Several diagnostic tools have been proposed, which are used in the daily practice and for research purposes, such as the M-ANNHEIM, Lüneburg, and Büchler diagnostic tools.[3–5] These diagnostic tools, however, have marked differences in the individual diagnostic criteria they are composed of (ie, clinical, morphological, and functional criteria). No studies that have compared the various diagnostic tools have been performed.

In this study, we evaluated and compared 3 widely used diagnostic tools for CP, ie, M-ANNHEIM, Büchler, and Lüneburg diagnostic tools. The aims of this study were to (1) evaluate and compare the individual criteria included in these tools; (2) to evaluate the effect of changes in these individual criteria on the diagnostic performance; (3) to evaluate the diagnostic agreement of the tools; and (4) to estimate their sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

MATERIALS AND METHODS

Study Design

We performed a cross-sectional analysis of a prospectively collected multicenter patient cohort including 669 patients with a first episode of acute pancreatitis (AP). These patients were retrospectively followed-up for a median period of 5 years, through medical records and patient's questionnaires, as described previously.[6,7] Variables needed to diagnose CP using M-ANNHEIM, Büchler, and the Lüneburg criteria were collected and scored.

Diagnostic Tools

The M-ANNHEIM Diagnostic Tool

This tool was developed in 2007 and distinguishes patients with definite CP and patients that are likely to have CP (ie, probable CP). Chronic pancreatitis according to the M-ANNHEIM tool is defined by the following criteria (Supplemental Table 1, http://links.lww.com/MPA/A607)[3]: a typical clinical history of CP (such as recurrent pancreatitis [RP] or abdominal pain) and 1 or more of the following additional criteria:

–Definite CP: pancreatic calcifications, or moderate or marked ductal lesions (according to the Cambridge classification), or marked and persistent exocrine insufficiency (pancreatic steatorrhea markedly reduced by enzyme supplementation) or typical histology of an adequate histological specimen.
–Probable CP: mild ductal alterations (according to the Cambridge classification), or recurrent or persistent pseudocysts, or pathological test of pancreatic exocrine function (such as fecal elastase-1 test, secretin test, and secretin-pancreozymin test), or endocrine insufficiency (abnormal glucose tolerance test).

The Büchler Diagnostic Tool

The Büchler criteria were developed in 2009 in Heidelberg as part of an attempt to introduce a simplified classification for different stages of CP.[4] For the diagnosis of CP, patients need to have at least 1 clinical criteria (ie, pain, RP, steatorrhea, diabetes mellitus [DM] or complications of CP), accompanied by well-defined imaging abnormalities of the pancreatic duct or parenchymal changes, or abnormal direct pancreatic function tests (Supplemental Table 2, http://links.lww.com/MPA/A607).

From the *Department of Surgery, Academic Medical Center, Amsterdam; †Department of Surgery, St Antonius Hospital, Nieuwegein; and ‡Department of Surgery, University Medical Center Utrecht, Utrecht, The Netherlands.
Received for publication October 10, 2016; accepted July 10, 2017.
Address correspondence to: Yama Issa, MD, PhD, Academic Medical Center, Department of Surgery (suite G4-178), Meibergdreef 9, 1105AZ, Amsterdam, The Netherlands (e-mail: y.issa@amc.nl).
Authors M.A.B. and M.G.B. declare relationships with the following companies: Baxter, Acelity/KCI, LifeCell, Ipsen, Mylan, Johnson & Johnson, and Bard. The authors of this article declare no relationships with companies whose products or services are the subject matter of the article.
Supplemental digital contents are available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.pancreasjournal.com).

DOI: 10.1097/MPA.0000000000000903

The Lüneburg Diagnostic Tool

The Lüneburg tool assigns points to certain histological, morphological, and functional characteristics.[5] A score of 4 or more establishes the diagnosis of CP. Points are assigned as follows: 4 points for pancreatic calcifications, typical histological changes, postmortem diagnosis of CP, and/or intraoperative findings of CP; 3 points for abnormal imaging procedures (computed tomography [CT], magnetic resonance cholangiopancreatography, ultrasonography, endoscopic retrograde cholangiopancreatography); 3 to 2 points for exocrine pancreatic function tests (secretin pancreozymin, pancreolauryl, fecal chymotrypsin, fecal elastase-1); and 1 point for steatorrhea (Supplemental Table 3, http://links.lww.com/MPA/A607).

The Lüneburg tool resembles the Mayo Clinic diagnostic tool, which is used more often in the United States, but the Lüneburg tool also considers the findings of indirect pancreatic function tests and findings on the ultrasound and diagnostic procedures for the diagnosis of CP.[8] Therefore, it was decided that separate assessment of the Mayo Clinic tool would not add useful information.

Clinical Diagnosis

Chronic pancreatitis is clearly stated as a diagnosis by a treating physician (specialist diagnosis) in the medical charts.

Data Analysis

Evaluate and Compare the Individual Components of the Included Diagnostic Tools

We first analyzed the incidence of individual components of each diagnostic tool, and we evaluated the discrepancies between the tools.

Descriptive Analysis of Discrepancies Between Different Diagnostic Tools

We examined and described the reasons for discrepancies between the diagnostic tools by evaluating the patients diagnosed with CP by any diagnostic tool and the reasons for nonuniform diagnosis.

Examine the Influence of the Components on Diagnosis of CP

We analyzed the influence of changes in individual components on the total number of diagnosed patients with CP for each diagnostic tool. This was done by adding (if not part of the original tool) or removing (if part of the tool) of each individual component, and evaluating how this altered the total number of patients diagnosed with CP. For the Lüneburg tool, points had to be assigned to criteria that were not part of the original tool. This was done as follows: 1 point for abdominal pain or RP, and 2 points for DM, mild to marked ductal lesions, enlargement of the pancreas, heterogeneous reflectivity, pancreatic pseudocysts, or other complications.

The Agreement Between Diagnostic Tools for the Diagnosis of CP

We calculated the strength of the agreement between the different diagnostic tools by using the κ coefficient.[8] Outcomes were described as follows: poor, less than 0.00; slight, 0.01 to 0.20; fair, 0.21 to 0.40; moderate, 0.41 to 0.60; substantial, 0.61 to 0.80; and (almost) perfect, 0.81 to 1.00.[9]

Diagnostic Accuracy Estimates

Because no criterion standard for the diagnosis of CP exists, it is not possible to truly evaluate diagnostic accuracy for the included diagnostic tools. Therefore, we used Bayesian latent-class analysis to estimate the sensitivity and specificity of the diagnostic tools as well as the prevalence of disease.[10] This method fits a distribution that minimizes the error based on the results of the 3 diagnostic tools in the same population. The model evaluating 3 tests in one population (the 2-dependent and 1-independent version) as described by Branscum et al[11] was used. We assumed the M-ANNHEIM criteria to be independent of the other 2 criteria. The model was fitted using Markov Chain Monte Carlo estimation using the WinBUGS software version 1.4.3 (The BUGS Project, London, UK). For the analyses presented, posterior inferences were based on 100,000 iterations after a burn-in of 10,000 iterations were discarded. Convergence was assessed by running 5 chains from dispersed starting values (prevalence from 5% to 15% and sensitivity and specificity from 75% to 95%).[12] We calculated the PPVs and NPVs using the following formulas: PPV = (sensitivity * prevalence)/sensitivity * prevalence + (1 − specificity) * (1 − prevalence) and NPV = specificity * (1 − prevalence)/(1 − sensitivity) * prevalence + specificity * (1 − prevalence).

Statistical Analysis

Data are presented as numbers and percentages, mean (standard deviation [SD]) or median (interquartile range), as appropriate. When comparing 2 diagnostic criteria within the cohort, a McNemar test for paired data was used. Statistical significance threshold was considered to be $P < 0.05$. In case of multiple testing, P values were corrected for multiple testing with the Benjamini-Hochberg method when appropriate.[13] Data were analyzed using IBM SPSS Statistics version 20.0 (Armonk, NY), unless otherwise specified.

RESULTS

Evaluate and Compare the Individual Components of the Included Diagnostic Tools

The M-ANNHEIM, Büchler, and Lüneburg diagnostic tools were applied to a total of 669 patients with AP. After a median period of 57 months (interquartile range, 42–70), CP was diagnosed in 50 (7%), 59 (9%), and 61 (9%) patients by the M-ANNHEIM definite, Lüneburg, and Büchler tools, respectively. When M-ANNHEIM probable criteria were applied, 60 (9%) were identified as likely to have CP. The treating physician diagnosed CP in 46 patients (7%) (Table 1).

Descriptive Analysis of Discrepancies Between Different Diagnostic Tools

A total of 77 patients were diagnosed with CP out of the total cohort of 669 patients (11.5%) with AP by any of the diagnostic tools, and a consensus among tools was found in 38 (49%). Patients with abdominal pain and significant steatorrhea but without morphological abnormalities on imaging were exclusively diagnosed with CP according to the M-ANNHEIM diagnostic tool. This occurred in 4 patients (5%). For the Lüneburg diagnostic tool: 11 patients (14%) had isolated (minimal) pancreatic calcifications without any other morphological, functional, or clinical signs of CP. These patients were thus exclusively diagnosed as having CP using the Lüneburg criteria, because, for both the M-ANNHEIM and the Büchler tools, at least 1 clinical criterion is needed. Regarding patients exclusively diagnosed by the Büchler tool: almost all

TABLE 1. Incidence of the Individual Variables for the Diagnosis of CP According to Different Diagnostic Tools and Treating Physician

Criteria	Specialist Diagnosis, n (%)	M-ANNHEIM Definite, n (%)	M-ANNHEIM Probable, n (%)	Lüneburg, n (%)	Büchler, n (%)
No. patients	46 (7)	50 (7)	60 (9)	59 (9)	61 (9)
Pain					
RP	38 (83)	40 (80)	50 (83)		39 (64)
Abdominal pain	33 (72)	38 (76)	39 (65)		37 (61)
Pancreatic function					
DM	16 (35)		38 (63)		26 (43)
Steatorrhea	9 (20)	18 (36)*		14 (24)	16 (21)
Fecal elastase-1 level	6/12 (13)		7/14 (12)†	7/12 (12)†	
Pancreatic parenchyma					
General/focal enlargement of gland	6 (13)				7 (12)
Heterogeneous reflectivity	18 (39)				29 (48)
Pancreatic calcifications	16 (35)	21 (42)		41 (69)	30 (49)
Pancreatic duct					
Mild ductal lesions	13 (28)		15 (25)		18 (30)
Moderate or marked ductal lesions	25 (54)	31 (62)			32 (53)
Complications					
Pancreatic pseudocysts	25 (54)		31 (52)		34 (56)‡
Complications§	9 (20)				10 (16)
Morphology					
Abnormal imaging‖	30/90 (65)			38/90 (64)	
Histology	3/7 (7)	3/5 (6)		3/8 (5)	
Intraoperative findings of CP	4/6 (9)			4/9 (7)	
Postmortem diagnose of CP	0/1 (0)			0/1 (0)	

*Markedly reduced by enzyme supplementation.

†Secretin pancreozymin, pancreolauryl, fecal chymotrypsin, or fecal elastase-1.

‡With clinical signs.

§Duodenal, vascular, or bile duct obstruction/stenosis, pancreatic fistula, ascites, or other rare complications.

‖According to Cambridge criteria.

7 (9%) of the patients had DM, mild ductal lesions, pseudocysts, or heterogeneous reflectivity of the pancreas (see Table 2).

Influence of Individual Components on Diagnosis of CP

M-ANNHEIM Definite

Patients had either abdominal pain (20%) or RP (24%), or both (56%); the majority, in combination with moderate to marked ductal lesions (26%), exocrine insufficiency (18%), calcifications (14%), or a combination of moderate to marked ductal lesions and calcifications (18%). A significant reduction in the number of diagnosis of CP was seen when abdominal pain (20%, $P = 0.018$), RP (24%, $P < 0.001$), moderate or marked ductal lesions (26%, $P < 0.001$), or steatorrhea (18%, $P = 0.021$) were removed as individual criteria from the diagnostic. Adding DM as criteria to the M-ANNHEIM definite led to a significant rise of the total number of diagnosis of CP (36%, $P < 0.001$) (see Table 3).

M-ANNHEIM Probable

All patients had either abdominal pain (17%) or RP (35%), or both (48%), especially in combination with DM (43%) or with pancreatic pseudocysts (23%). Removing criteria as pancreatic pseudocysts or DM reduced the number of diagnosis with 27% ($P < 0.001$) and 43% ($P = 0.013$), respectively. Adding the criteria "abnormal findings on imaging" according to the Cambridge classification resulted in a significant increase of the number of diagnosis (18%, $P = 0.007$).

Lüneburg

Most diagnoses of CP were based on calcifications alone (29%), abnormal imaging according to the Cambridge classification (15%), or a combination of calcifications and abnormal imaging (24%). Adding abdominal pain or RP led to significant rise of the number diagnosis of CP by 14% and 22%, respectively. Removing pancreatic calcifications or abnormal findings on imaging according to Cambridge classification lead to a reducing of diagnosis by 51% and 24%, respectively ($P < 0.001$).

Büchler

For the clinical criteria, most of the patients had either abdominal pain (14%) or RP (17%), or both (48%). In 11% of cases, DM was the only clinical criteria. Regarding the additional criteria, the most prevalent were as follows: mild to marked ductal lesion (64%), pancreatic pseudocysts (56%), calcifications (49%), and heterogeneous reflectivity of the pancreas (48%). Interestingly, in 5% of the diagnosis, the presence of pancreatic cysts was enough for the diagnosis of CP. Removing or adding of individual criteria did not lead to significant changes in the number of diagnosis of CP.

TABLE 2. Patients Diagnosed With CP by Any Diagnostic Tool and the Reasons for Nonuniform Diagnosis

No. Patients	M-ANNHEIM	Büchler	Lüneburg	Reasons for Diagnosis of CP in a Tool
4	X			All had abdominal pain or RP and significant steatorrhea. None had imaging abnormalities.
7		X		– 2 Patients had abdominal pain with pancreatic cysts and heterogeneous reflectivity of the pancreas. – 3 Patients had pseudocysts with clinical signs with either pancreatic cysts and/or heterogeneous reflectivity of the pancreas. – 1 Patient had DM and a pancreatic pseudocyst on imaging. – 1 Patient had abdominal pain and a mild dilatation (<4 mm) of pancreatic duct.
11			X	All had isolated (minimal) pancreatic calcifications (of whom 1 patient had also pancreatic pseudocysts) without clinical signs.
7	X	X		All 7 patients had abdominal pain with either moderate to marked PD lesions, pseudocysts, and/or steatorrhea.
1	X		X	1 Patient with abdominal pain and RP with typical histology of CP
9		X	X	All 9 patients had calcifications in combination with DM (n = 7), abnormal imaging (n = 5), mild to marked PD lesions (n = 3), steatorrhea (n = 2), and/or heterogeneous reflectivity of the pancreas (n = 1). None had abdominal pain or RP.
38	X	X	X	CP according to all criteria

A total of 77 (11.5%) of 669 patients were diagnosed with CP by any of the diagnostic tools.

TABLE 3. The Effect of Adding or Removing an Individual Variable on the Outcome of Different Scoring Systems for CP

Variable	Δ MD, n (%)	Δ MP, n (%)	Δ L, n (%)	Δ B, n (%)
CP diagnosis, n	n = 50	n = 60	n = 59	n = 61
Pain				
RP	38 (−24)*	39 (−35)	72 (+22)†	57 (−7)
Abdominal pain	40 (−20)‡	50 (−17)	67 (+14)§	60 (−2)
Pancreatic function				
DM	68 (+36)*	34 (−43)‖	73 (+24)	55 (−10)
Steatorrhea	41 (−18)¶	63 (+5)	57 (−3)	61 (−0)
Fecal elastase-1 level	50 (+0)	60 (−0)	59 (−0)	62 (+2)
Morphology				
Abnormal imaging	58 (+16)	71 (+18)#	45 (−24)*	68 (+11)
Pancreatic calcifications	43 (−14)	60 (+0)	29 (−51)*	54 (−11)
Mild ductal lesions	51 (+2)	58 (−3)	59 (+0)	60 (−2)
Moderate or marked ductal lesions	37 (−26)*	64 (+7)	59 (+0)	60 (−2)
General/focal enlargement of gland	50 (+0)	61 (+2)	64 (+8)	61 (−0)
Heterogeneous reflectivity	52 (+4)	66 (+10)	73 (+24)	59 (−3)
Local complications				
Recurrent or persistent pseudocysts	52 (+4)	44 (−27)*	62 (+5)	61 (+0)
Complications	57 (+12)	66 (+10)	62 (+5)	61 (−0)
Histology	49 (−2)	60 (+0)	59 (−0)	61 (+0)
Intraoperative findings of CP	50 (+0)	60 (+0)	59 (−0)	61 (+0)
Postmortem diagnose of CP	50 (+0)	60 (+0)	59 (−0)	61 (+0)

Δ is change in tool after adding or removing of a variable.

McNemar test with Benjamini-Hochberg correction for multiple testing:

*$P < 0.001$.

†$P = 0.016$.

‡$P = 0.018$.

§$P = 0.044$.

‖$P = 0.013$.

¶$P = 0.021$.

#$P = 0.007$.

MD indicates M-ANNHEIM definite; MP, M-ANNHEIM probable; B, Büchler; L, Lüneburg.

TABLE 4. Agreement of the Diagnosis of CP by Different Diagnostic Tools

	Observed Agreement	κ (95% CI)	Strength of Agreement
Diagnostic tools			
M-ANNHEIM D vs Büchler	0.97	0.79 (0.70–0.88)	Substantial
Büchler vs Lüneburg	0.96	0.76 (0.67–0.88)	Substantial
M-ANNHEIM D vs Lüneburg	0.95	0.69 (0.59–0.79)	Substantial
Mean (SD)	0.96 (0.01)	0.75 (0.0)	Substantial
Diagnostic tools vs M-ANNHEIM probable			
M-ANNHEIM P vs M-ANNHEIM D	0.96	0.70 (0.60–0.80)	Substantial
M-ANNHEIM P vs Büchler	0.94	0.62 (0.51–0.73)	Substantial
M-ANNHEIM P vs Lüneburg	0.91	0.46 (0.34–0.58)	Moderate
Mean (SD)	0.94 (0.02)	0.60 (0.11)	Moderate
Diagnostic tools vs physician			
M-ANNHEIM P vs by physician	0.94	0.53 (0.41–0.65)	Moderate
M-ANNHEIM D vs by physician	0.96	0.71 (0.60–0.82)	Substantial
Büchler vs by physician	0.94	0.61 (0.50–0.72)	Substantial
Lüneburg vs by physician	0.93	0.52 (0.40–0.64)	Moderate
Mean (SD)	0.94 (0.01)	0.60 (0.09)	Moderate
Overall mean (SD)	0.95 (0.02)	0.65 (0.11)	Substantial

Observed agreement was calculated as the number of patients identified with the diagnosis by the different tools as present or absent divided by a total of 669 diagnoses.

CI indicates confidence interval.

The Agreement of the Diagnosis of CP by Different Diagnostic Criteria

The overall agreement between the M-ANNHEIM, Büchler, and Lüneburg diagnostic tools was substantial (κ = 0.75). The highest agreement was between M-ANNHEIM and Büchler (κ = 0.79) and lowest between the M-ANNHEIM and Lüneburg (κ = 0.69). Interestingly, the overall agreement between the diagnostic tools and the diagnosis made by the physician was moderate (κ = 0.60), ranging from 0.52 between Lüneburg and physician to 0.71 between M-ANNHEIM and physician (see Table 4).

Diagnostic Accuracy Estimates

With 94%, the Büchler diagnostic tool had the highest estimated sensitivity, followed by the M-ANNHEIM tool (87%), and finally the Lüneburg tool (81%). Moreover, the PPV ranged from 92% (M-ANNHEIM) to 72% (Lüneburg). The specificity ranges from 97% to 99%, probably because of the low prevalence rate of CP (8%). See Table 5.

DISCUSSION

In this study, we compared 3 of the most commonly used diagnostic tools for CP in a large cohort with patients with suspected CP. This offers a unique opportunity to evaluate the characteristics of these tools in clinical practice, including potential strengths and weaknesses. Several important findings were observed. First, we were able to find the differences between tools that resulted in the largest discrepancies in the diagnosis of CP. These were mainly the inclusion criteria of clinical symptoms (ie, abdominal pain or RP), DM, and certain morphological complications (eg, enlarged glands, pseudocysts, and heterogeneous reflectivity). Differences in other criteria, such as histology and mild PD lesions, did not result in large differences. Second, despite differences, the agreement between the various tools was substantial. On the other hand, the agreement between diagnosis by physician and different tools was much lower. This emphasizes the importance of a methodological approach to the diagnosis of a complex disease such as CP. Finally, using a Bayesian approach we were able to estimate the diagnostic accuracy of the different tools. This provides insights in how the different tools compare but should be interpreted with caution owing to lack of a reference test.

The 3 studied tools showed substantial differences when compared with each other. Of the 77 patients diagnosed with CP by any tool, all 3 tools agreed only on 38 of them (49% consensus rate). Certain important patterns were identified when

TABLE 5. Diagnostic Sensitivity, Specificity, PPV and NPV, and the Prevalence Rates of the Different Diagnostic Tools

	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV
Büchler	94% (91%–96%)	98% (98%–99%)	82%	99%
Lüneburg	81% (77%–85%)	97% (97%–98%)	72%	98%
M-ANNHEIM	87% (82%–92%)	99% (99%–100%)	92%	99%
Prevalence (95% CI)		8.1% (7.3%–8.9%)		

Data expressed in median (interquartile range).

CI indicates confidence interval.

examining the differences between the various diagnostic tools. For example, we found that patients with abdominal pain and marked steatorrhea that improved with pancreatic enzymes without any imaging abnormalities were diagnosed with definite CP according to the M-ANNHEIM tool, but not according to the other tools. Similarly, patients with clinical symptoms and either DM, pseudocysts, or mild morphological changes (eg, heterogeneous reflectivity) were exclusively diagnosed by the Büchler tool. The effect of these differences proved to be substantial. Adding DM, for example, as a criterion to the M-ANNHEIM tool, showed a significant effect (increase of 36%) on the total number of diagnosed patients. For the Büchler tool, 5% of the patients with CP were diagnosed based on the presence of pancreatic pseudocysts as the only morphological criteria. The Lüneburg tool consists of a point system and uses different forms of exocrine function tests, which are not used often nowadays. Two important features of this tool were observed. First, patients with isolated (minimal) pancreatic calcifications after a single episode of AP without any clinical, functional, or morphological signs of CP were diagnosed with CP. Second, patients with typical clinical symptoms and abnormal morphology (other than calcification) on 1 imaging modality were not diagnosed for CP if they had no exocrine pancreatic insufficiency, histological examination, or intraoperative findings.

The overall agreement between the diagnostic tools was substantial with an average κ of 0.75. However, the agreement between diagnosis by physician and the different tools was much lower. The relatively high agreement between the tools despite the differences outlined previously probably reflects the systematic methodology used to come to final decision in such tools. This methodology is frequently lacking in clinical practice, in which many physicians diagnose CP without use of formal criteria. These findings once again illustrate the importance of the use of a well-designed diagnostic tool in complex diseases such as CP.

Using Bayesian methods, we were able to estimate diagnostic accuracy parameters for the different tools. The Büchler tool had the highest sensitivity (ie, lowest number false negatives) compared with the other tools. This can be explained by the extensiveness of this tool, which allows for far more choices of clinical and morphology criteria to diagnose CP. The number of potential false positives was moderate (7 patients exclusively diagnosed by Büchler, compared with 4 by M-ANNHEIM and 11 Lüneburg), which resulted in a reasonable estimated PPV (82%). In line with this argumentation, the M-ANNHEIM tool had the highest positive predicted value (92%). The Lüneburg tool had the lowest estimates of sensitivity and specificity. This is probably owing to the 2 features of this tool explained previously, which were responsible for a large number of potential false positives and negatives. It should be noted that the differences in specificity were small in this cohort, which is explained by the (relatively) low prevalence (8.1%) of CP in the study population.

There is a lack of consensus regarding the diagnosis of CP. Pancreatic histology is seen as a criterion standard, but is rarely available. Currently, CP diagnosis is mainly based on imaging studies to detect morphological abnormalities. Although direct exocrine pancreatic function tests are well represented in the diagnostic tools, at present, they are rarely, if ever performed, limiting the role for exocrine pancreatic function in the diagnosis of CP. Alternative (indirect) tests, such as fecal-elastase or acid-steatocrit tests might offer an important addition for the diagnosis of CP in cases with inconclusive morphological findings. When examining the different tools, both Büchler and M-ANNHEIM offer reasonable alternative for clinical practice. The M-ANNHEIM tool is more concise and thus potentially easier for clinical practice. On the other hand, the Büchler tool offers more comprehensive set of criteria, thus potentially allowing for diagnosis of patients with less typical presentation of CP. The 2 main topics in need for further research to differentiate between these 2 tools are whether DM can be considered a reliable clinical symptom of CP and whether certain morphological changes (eg, enlargement of the gland, cysts, and heterogeneous reflectivity) are diagnostic of CP in symptomatic patients. As for the Lüneburg tool, we feel that it is outdated in its current form for 2 major reasons. First, with advances in imaging (especially high-resolution CT), even minute calcifications can be seen after a single episode of AP. In otherwise asymptomatic patients, this should not lead to the diagnosis of CP. Second, this score does not take clinical symptoms into consideration, requiring some form of exocrine pancreatic insufficiency instead. Thus, in patients with typical symptoms and typical morphological imaging abnormalities, the diagnosis of CP might still not be established.

Especially difficult to diagnose are patients in an early stage of CP, with minimal morphological abnormalities on imaging. Endoscopic ultrasound (EUS) and perhaps secretin enhanced magnetic resonance imaging (sMRI) may provide an add-on value when MRI and CT are inconclusive, for detecting these minimal structural changes.[14–18] For a comprehensive diagnostic tool, one cannot base the diagnosis on clinical symptoms alone. For example, some patients experienced continuous or recurrent abdominal pain, thought to be of pancreatic origin, but do not show any morphologic or functional abnormalities. However, there are also patients with morphologic abnormalities on imaging that have never experience pain (primary painless pancreatitis). Furthermore, the diagnosis of different rare entities of CP, such as groove pancreatitis or autoimmune pancreatitis, makes it more complex to make an overall diagnostic tool. For example, the diagnosis off groove pancreatitis is based on the duodeno-pancreatic groove, with the presence of cysts surrounded by inflammation and fibrosis.[19–22] The diagnosis of autoimmune pancreatitis is usually based on a combination of imaging findings, serology, sometimes with other organ involvement, histology, and response to corticosteroids.[23,24]

Strengths of this study include the high number of patients from both academic and nonacademic hospitals. Our results are therefore probably generalizable to the general patient population with CP. Another strength was the presence of a detailed description of the clinical characteristics of the patients and imaging data, also during follow-up. The study also has limitations. First, the data were partly collected retrospectively. Second, the population in which the diagnostic tools were analyzed were patients with a first episode of AP. Some could argue that this could lead to a low prevalence of CP, and a population of patients with suspected CP would be more preferable to analyze. However, these patients are at risk of developing CP (especially with high incidence of smoking and alcohol consumption), which makes this population relevant for this study.

Substantial differences exist between diagnostic tools for CP, leading to differences in the number of patients diagnosed with CP. The Büchler tool had the highest sensitivity (94%). Differences in whether or not CP is diagnosed are mainly attributed to whether or not clinical symptoms (ie, abdominal pain or RP), endocrine insufficiency, and certain morphological complications (eg, enlarged glands, pseudocysts, and heterogeneous reflectivity) are included in the tools. Differences in other criteria, such as histology and mild PD lesions, did not result in large differences. Although no optimal diagnostic tool for CP exists and designing such tool is challenging, this study creates awareness of the differences among tools leading to different numbers of CP diagnosis and offers guidance for future research.

REFERENCES

1. Delhaye M, Van Steenbergen W, Cesmeli E, et al. Belgian consensus on chronic pancreatitis in adults and children: statements on diagnosis and nutritional, medical, and surgical treatment. *Acta Gastroenterol Belg*. 2014;77:47–65.
2. Conwell DL, Lee LS, Yadav D, et al. American Pancreatic Association Practice Guidelines in Chronic Pancreatitis: evidence-based report on diagnostic guidelines. *Pancreas*. 2014;43:1143–1162.
3. Schneider A, Löhr JM, Singer MV. The M-ANNHEIM classification of chronic pancreatitis: introduction of a unifying classification system based on a review of previous classifications of the disease. *J Gastroenterol*. 2007;42:101–119.
4. Büchler MW, Martignoni ME, Friess H, et al. A proposal for a new clinical classification of chronic pancreatitis. *BMC Gastroenterol*. 2009;9:93.
5. Lankisch PG, Breuer N, Bruns A, et al. Natural history of acute pancreatitis: a long-term population-based study. *Am J Gastroenterol*. 2009;104:2797–2805; quiz 2806.
6. Ahmed Ali U, Issa Y, Hagenaars JC, et al. Risk of recurrent pancreatitis and progression to chronic pancreatitis after a first episode of acute pancreatitis. *Clin Gastroenterol Hepatol*. 2016;14:738–746.
7. Besselink MG, van Santvoort HC, Boermeester MA, et al. Timing and impact of infections in acute pancreatitis. *Br J Surg*. 2009;96:267–273.
8. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37–46.
9. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*. 1977;33:363–374.
10. Rutjes AW, Reitsma JB, Coomarasamy A, et al. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess*. 2007;11:iii, ix-51.
11. Branscum AJ, Gardner IA, Johnson WO. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Prev Vet Med*. 2005; 68:145–163.
12. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Science*. 1992;7:457–472.
13. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol*. 1995;57:289–300.
14. Matos C, Metens T, Devière J, et al. Pancreatic duct: morphologic and functional evaluation with dynamic MR pancreatography after secretin stimulation. *Radiology*. 1997;203:435–441.
15. Nicaise N, Pellet O, Metens T, et al. Magnetic resonance cholangiopancreatography: interest of IV secretin administration in the evaluation of pancreatic ducts. *Eur Radiol*. 1998;8:16–22.
16. Cappeliez O, Delhaye M, Devière J, et al. Chronic pancreatitis: evaluation of pancreatic exocrine function with MR pancreatography after secretin stimulation. *Radiology*. 2000;215:358–364.
17. Sherman S, Freeman ML, Tarnasky PR, et al. Administration of secretin (RG1068) increases the sensitivity of detection of duct abnormalities by magnetic resonance cholangiopancreatography in patients with pancreatitis. *Gastroenterology*. 2014;147:646–654.e2.
18. Trikudanathan G, Walker SP, Munigala S, et al. Diagnostic performance of contrast-enhanced MRI with secretin-stimulated MRCP for non-calcific chronic pancreatitis: a comparison with histopathology. *Am J Gastroenterol*. 2015;110:1598–1606.
19. Rebours V, Lévy P, Vullierme MP, et al. Clinical and morphological features of duodenal cystic dystrophy in heterotopic pancreas. *Am J Gastroenterol*. 2007;102:871–879.
20. Fléjou JF. Paraduodenal pancreatitis: a new unifying term and its morphological characteristics. *Diagn Histopathol*. 2012;18:31–36.
21. Adsay NV, Zamboni G. Paraduodenal pancreatitis: a clinico-pathologically distinct entity unifying "cystic dystrophy of heterotopic pancreas", "para-duodenal wall cyst", and "groove pancreatitis". *Semin Diagn Pathol*. 2004;21:247–254.
22. Triantopoulou C, Dervenis C, Giannakou N, et al. Groove pancreatitis: a diagnostic challenge. *Eur Radiol*. 2009;19:1736–1743.
23. Chari ST, Kloeppel G, Zhang L, et al. Histopathologic and clinical subtypes of autoimmune pancreatitis: the Honolulu consensus document. *Pancreas*. 2010;39:549–554.
24. Kamisawa T, Chari ST, Lerch MM, et al. Recent advances in autoimmune pancreatitis: type 1 and type 2. *Gut*. 2013;62:1373–1380.