UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Advances in epidemiology and management of chronic pancreatitis

Ahmed Ali, U.

Publication date
2019
Document Version
Other version
License
Other

Link to publication

Citation for published version (APA):
Ahmed Ali, U. (2019). *Advances in epidemiology and management of chronic pancreatitis*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:09 Mar 2023

Y. Issa[1], H.C. van Santvoort[2], S. van Dieren[1], M.G. Besselink[1], M.A. Boermeester[1], U. Ahmed Ali[1,3]

[1] Department of Surgery, Academic Medical Center, Amsterdam, the Netherlands
[2] Department of Surgery, St Antonius Hospital, Nieuwegein, the Netherlands
[3] Department of Surgery, University Medical Center Utrecht, Utrecht, the Netherlands

2 DIAGNOSING CHRONIC PANCREATITIS: COMPARISON AND EVALUATION OF DIFFERENT DIAGNOSTIC TOOLS

Pancraes, 2017.

ABSTRACT

Background

The aim of the study was to compare the Mannheim, Büchler and Lüneburg diagnostic tools for chronic pancreatitis (CP).

Methods

A cross-sectional analysis was performed of the development of CP in a prospectively collected multi-centre cohort including 669 patients after a first episode of acute pancreatitis (AP). We compared the individual components of the Mannheim, 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 (IQR 42-70) months were included. CP was diagnosed in 50 patients (7%), 59 patients (9%) and 61 patients (9%) by the Mannheim, 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 Mannheim (87%) and finally the Lüneburg tool (81%).

Conclusion

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

INTRODUCTION

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, unnecessary resource utilisation 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 Mannheim, Lüneburg, and Büchler diagnostic tools.[3-5] These diagnostic tools, however, have marked differences in the individual diagnostic criteria they are composed of (i.e. clinical, morphological and functional criteria). No studies have been performed that have compared the various diagnostic tools.

In this study we evaluated and compared three widely used diagnostic tools for CP, i.e. Mannheim, 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; 4) to estimate their sensitivity, specificity, positive predictive value and negative predictive value.

METHODS

Study design

We performed a cross-sectional analysis of a prospectively collected multi-center 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 Mannheim, Büchler and the Lüneburg criteria were collected and scored.

Diagnostic tools

The Mannheim diagnostic tool

This tool was developed in 2007 and distinguishes patients with definite CP and patients that are likely to have CP (i.e. probable CP). CP according to the Manhnheim tool is defined by the following criteria (Supplemental Table A1):[3] a typical clinical history of CP (such as recurrent pancreatitis (RP) or abdominal pain) and one 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, 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 (i.e. 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 A2).

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: Four points for pancreatic calcifications, typical histological changes, post-mortem diagnosis of CP and/or intra-operative findings of CP. Three points for abnormal imaging procedures (CT, MRCP, US, ERCP) and 3 to 2 points for exocrine pancreatic function tests (secretin pancreozymin, pancreolauryl, fecal chymotrypsin, fecal elastase-1) and 1 point for steatorrhea (Supplemental Table A3).

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

CP 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 analysed 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 non-uniform diagnosis.

Examine the influence of the components on diagnosis of CP

We analysed 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 recurrent pancreatitis, 2 points for diabetes mellitus, 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 kappa (κ) coefficient.[9] Outcomes were described as: poor <0.00, slight 0.01-0.20, fair 0.21-0.40, moderate 0.41-0.60, substantial 0.61-0.80, and (almost) perfect 0.81-1.00.[10]

Diagnostic accuracy estimates

Since no golden-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.[11] This method fits a distribution that minimizes the error based on the results of the three diagnostic tools in the same population. The model evaluating three tests in one population (the 2-dependent and 1-independent version) as described by Branscum et al. was used.[12] We assumed the M-ANNHEIM criteria to be independent of the other two criteria. The model was fitted using Markov Chain Monte Carlo estimation using the WinBUGS software. 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%).[13] We calculated the positive and negative predictive values using sensitivity, specificity and prevalence as usual.

Statistical analysis

Data are presented as numbers and percentages, mean [± standard deviation (SD)] or median [interquartile range (IQR)], as appropriate. When comparing two 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.[14] Data were analysed using IBM SPSS Statistics version 20.0, unless otherwise specified.

RESULTS

Evaluate and compare the individual components of the included diagnostic tools

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

Table 1. Incidence of the individual variables for the diagnosis of chronic pancreatitis according to different diagnostic tools and treating physician

Criteria	Specialist diagnosis N (%)	Mannheim Definite N (%)	Mannheim Probable N (%)	Lüneburg N (%)	Büchler N (%)
Number of patients	46 (7)	50 (7)	60 (9)	59 (9)	61 (9)
Pain					
- Recurrent pancreatitis	38 (83)	40 (80)	50 (83)		39 (64)
- Abdominal pain	33 (72)	38 (76)	39 (65)		37 (61)
Pancreatic Function					
- Diabetes mellitus	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)

Table 1. Incidence of the individual variables for the diagnosis of chronic pancreatitis according to different diagnostic tools and treating physician (*continued*)

Criteria	Specialist diagnosis N (%)	Mannheim Definite N (%)	Mannheim Probable N (%)	Lüneburg N (%)	Büchler N (%)
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)	
- Intra-operative findings of CP	4/6 (9%)			4/9 (7)	
- Post mortem diagnose of CP	0/1 (0%)			0/1 (0)	

MD=Mannheim definite, MP= Mannheim probable, B=Büchler, L=Lüneburg
* markedly reduced by enzyme supplementation, † secretin pancreozymin, pancreolauryl, fecal chymotrypsin, fecal elastase-1, ‡ with clinical signs, § duodenal, vascular or bile duct obstruction/stenosis, pancreatic fistula, ascites or other rare complications, ‖ according to Cambridge criteria,

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 with acute pancreatitis (11.5%) 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 the Mannheim diagnostic. This occurred in 4 (5%) patients. For the Lüneburg diagnostic tool: 11 (14%) patients 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, since for both the Mannheim and the Büchler tools at least 1 clinical criterion is needed. Regarding patients exclusively diagnosed by the Büchler tool: almost all 7 (9%) of the patients had DM, mild ductal lesions, pseudocysts or heterogeneous reflectivity of the pancreas (Table 2).

Table 2. Patients diagnosed with chronic pancreatitis by any diagnostic tool and the reasons for non-uniform diagnosis

Number of patients	Mannheim	Büchler	Lüneburg	Reasons for diagnosis of CP in a tool
4	X			All had abdominal pain or recurrent pancreatitis and significant steatorrhea. None had imaging abnormalities
7		X		- 2 patients had abdominal pain with pancreatic cysts and heterogeneous reflectivity of the pancreas - 3 patient had pseudocysts with clinical signs with either pancreatic cysts and/or heterogenous reflectively of the pancreas - 1 patient had DM and a pancreatic pseudocyst on imaging - 1 patient abdominal pain and a mild dilatation (<4mm) 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 recurrent pancreatitis 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 heterogenous reflectively of the pancreas (n=1). None had abdominal pain or recurrent pancreatitis
38	X	X	X	CP according to all criteria

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

Influence of individual components on diagnosis of CP

The effect of adding or removing an individual variable on the outcome of different scoring systems for CP is presented in Table 3. The effect on each of the scoring systems is discussed below.

Mannheim 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 Mannheim definite lead to a significant rise of the total number of diagnosis of CP (36%, P<0.001).

Table 3. The effect of adding or removing an individual variable on the outcome of different scoring systems for chronic pancreatitis

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				
- Recurrent pancreatitis	**38 (-24)***	39 (-35)	**72 (+22)**†	57 (-7)
- Abdominal pain	**40 (-20)**‡	50 (-17)	**67 (+14)§**	60 (-2)
Pancreatic Function				
- Diabetes mellitus	**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)
- Intra-operative findings of CP	50 (+0)	60 (+0)	59 (-0)	61 (+0)
- Post mortem diagnose of CP	50 (+0)	60 (+0)	59 (-0)	61 (+0)

Δ = Change in tool after adding or removing of a variable
MD=Mannheim Definite, MP= Mannheim Probable, B=Büchler, L=Lüneburg
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

Mannheim 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 lead 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 the majority 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: 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 a 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.

The agreement of the diagnosis of CP by different diagnostic criteria

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

Table 4. Agreement of the diagnosis of chronic pancreatitis by different diagnostic tools

	Observed agreement	Kappa (95% CI)	Strength of agreement
Diagnostic tools			
MannheimD vs Buchler	0.97	0.79 (0.70-0.88)	Substantial
Buchler vs Lüneburg	0.96	0.76 (0.67-0.88)	Substantial
MannheimD 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 Mannheim probable			
MannheimP vs MannheimD	0.96	0.70 (0.60-0.80)	Substantial
MannheimP vs Buchler	0.94	0.62 (0.51-0.73)	Substantial
MannheimP 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			
MannheimP vs By Physician	0.94	0.53 (0.41-0.65)	Moderate
MannheimD vs By Physician	0.96	0.71 (0.60-0.82)	Substantial
Buchler vs By Physician	0.94	0.61 (0.50-0.72)	Substantial

Table 4. Agreement of the diagnosis of chronic pancreatitis by different diagnostic tools (*continued*)

	Observed agreement	Kappa (95% CI)	Strength of agreement
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 diagnosis.

Diagnostic accuracy estimates

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

Table 5. Diagnostic sensitivity, specificity, positive and negative predicted value and the prevalence rates of the different diagnostic tools expressed in median (interquartile ranges)

	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%
Mannheim	87% (82% – 92%)	99% (99%-100%)	92%	99%
Prevalence (95% CI)		8.1% (7.3% - 8.9%)		

DISCUSSION

In this study, we compared three 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 (i.e. abdominal pain or recurrent pancreatitis), diabetes mellitus and certain morphological complications (e.g. enlarged glands, pseudocysts, heterogeneous reflectivity). Differences in other criteria, such as histology and mild PD lesions, did not result in large differences. Secondly, 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 due to lack of a reference test.

The three studied tools showed substantial differences when compared to each other. Of the 77 patients diagnosed with CP by any tool, all three tools agreed only on 38 of them (49% consensus rate). Certain important patters were identified when 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 Mannheim tool, but not according to the other tools. Similarly, patients with clinical symptoms and either DM, pseudocysts or mild morphological changes (e.g. heterogeneous reflectively) 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 Mannheim tool, showed a significant effect (increase of 36%) on the total number of diagnosed patients. For the Büchler tool, five percent 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 acute pancreatitis without any clinical, functional or morphological signs of CP, were diagnosed with CP. Secondly, patients with typical clinical symptoms and abnormal morphology (other than calcification) on one 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 Kappa 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 above, 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 (i.e. lowest number false negatives) compared to 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 to 4 by Mannheim and 11 Lüneburg), which resulted in a reasonable

estimated positive predictive value (82%). In line with this argumentation, the Mannheim tool had the highest positive predicted value (92%). The Lüneburg tool had the lowest estimates of sensitivity and specificity. This is probably due to the two features of this tool explained above, 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 gold standard, but is rarely available. Currently, CP diagnosis is mainly based on imaging studies to detect morphological abnormalities. While direct exocrine pancreatic function tests are well represented in the diagnostic tools, at present there 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 Mannheim offer reasonable alternative for clinical practice. The Mannheim 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 two main topics in need for further research to differentiate between these two tools are whether DM can be considered a reliable clinical symptom of CP, and whether certain morphological changes (e.g. 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 two major reasons. First, with advances in imaging (especially high resolution CT), even minute calcifications can be seen after a single episode of acute pancreatitis. In otherwise asymptomatic patients, this should not lead to the diagnosis of CP. Secondly, 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 computed tomography (CT) are inconclusive, for detecting these minimal structural changes.[15-19] 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 auto-immune 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.[20-23] The diagnosis of auto-immune pancreatitis is usually based on a combination of imaging findings, serology, sometimes with other organ involvement, histology, and response to corticosteroids.[24,25]

Strengths of this study include the high number of patients from both academic and non-academic 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 analysed were patients with a first episode of acute pancreatitis. 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 analyse. 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 (i.e. abdominal pain or recurrent pancreatitis), endocrine insufficiency, and certain morphological complications (e.g. enlarged glands, pseudocysts, heterogeneous reflectivity) are included in the tools. Differences in other criteria, such as histology and mild PD lesions, did not result in large differences. While 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, Pelckmans P, Putzeys V, Roeyen G, Berrevoet F, et al. Belgian consensus on chronic pancreatitis in adults and children: Statements on diagnosis and nutritional, medical, and surgical treatment. Acta gastro-enterologica Belgica. 2014;77:47-65.
2. Conwell DL, Lee LS, Yadav D, Longnecker DS, Miller FH, Mortele KJ, Levy MJ, et al. American pancreatic association practice guidelines in chronic pancreatitis: Evidence-based report on diagnostic guidelines. Pancreas. 2014;43:1143-62.
3. Schneider A, Lohr 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. Journal of gastroenterology. 2007;42:101-19.
4. Buchler MW, Martignoni ME, Friess H, Malfertheiner P. A proposal for a new clinical classification of chronic pancreatitis. BMC gastroenterology. 2009;9:93.
5. Lankisch PG, Breuer N, Bruns A, Weber-Dany B, Lowenfels AB, Maisonneuve P. Natural history of acute pancreatitis: A long-term population-based study. The American journal of gastroenterology. 2009;104:2797-805; quiz 806.
6. Ahmed Ali U, Issa Y, Hagenaars JC, Bakker OJ, van Goor H, Nieuwenhuijs VB, Bollen TL, et al. Risk of recurrent pancreatitis and progression to chronic pancreatitis after a first episode of acute pancreatitis. Clinical gastroenterology and hepatology: the official clinical practice journal of the American Gastroenterological Association. 2016;14: 738-46.
7. Besselink MG, van Santvoort HC, Boermeester MA, Nieuwenhuijs VB, van Goor H, Dejong CH, Schaapherder AF, et al. Timing and impact of infections in acute pancreatitis. The British journal of surgery. 2009;96:267-73.
8. Layer P, Yamamoto H, Kalthoff L, Clain JE, Bakken LJ, DiMagno EP. The different courses of early- and late-onset idiopathic and alcoholic chronic pancreatitis. Gastroenterology. 1994;107:1481-7.
9. Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 1960;20:37-46.
10. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics. 1977;33: 363-74.
11. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. Health technology assessment (Winchester, England). 2007; 11:iii, ix-51.
12. Branscum AJ, Gardner IA, Johnson WO. Estimation of diagnostic-test sensitivity and specificity through bayesian modeling. Preventive veterinary medicine. 2005;68: 145-63.
13. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical Science. 1992;7:457-72.
14. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological). 1995;57:289-300.
15. Matos C, Metens T, Deviere J, Nicaise N, Braude P, Van Yperen G, Cremer M, et al. Pancreatic duct: Morphologic and functional evaluation with dynamic mr pancreatography after secretin stimulation. Radiology. 1997;203:435-41.
16. Nicaise N, Pellet O, Metens T, Deviere J, Braude P, Struyven J, Matos C. Magnetic resonance cholangiopancreatography: Interest of iv secretin administration in the evaluation of pancreatic ducts. European radiology. 1998;8:16-22.

17. Cappeliez O, Delhaye M, Deviere J, Le Moine O, Metens T, Nicaise N, Cremer M, et al. Chronic pancreatitis: Evaluation of pancreatic exocrine function with mr pancreatography after secretin stimulation. Radiology. 2000;215:358-64.
18. Sherman S, Freeman ML, Tarnasky PR, Wilcox CM, Kulkarni A, Aisen AM, Jacoby D, 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: 646-54.e2.
19. Trikudanathan G, Walker SP, Munigala S, Spilseth B, Malli A, Han Y, Bellin M, et al. Diagnostic performance of contrast-enhanced mri with secretin-stimulated mrcp for non-calcific chronic pancreatitis: A comparison with histopathology. The American journal of gastroenterology. 2015;110:1598-606.
20. Rebours V, Levy P, Vullierme MP, Couvelard A, O'Toole D, Aubert A, Palazzo L, et al. Clinical and morphological features of duodenal cystic dystrophy in heterotopic pancreas. The American journal of gastroenterology. 2007;102:871-9.
21. Flejou J. Paraduodenal pancreatitis: A new unifying term and its morphological characteristics. Diagnostic Histopathology. 2012; 18:31-36.
22. Adsay NV, Zamboni G. Paraduodenal pancreatitis: A clinico-pathologically distinct entity unifying "cystic dystrophy of heterotopic pancreas", "para-duodenal wall cyst", and "groove pancreatitis". Seminars in diagnostic pathology. 2004;21:247-54.
23. Triantopoulou C, Dervenis C, Giannakou N, Papailiou J, Prassopoulos P. Groove pancreatitis: A diagnostic challenge. European radiology. 2009;19:1736-43.
24. Chari ST, Kloeppel G, Zhang L, Notohara K, Lerch MM, Shimosegawa T. Histopathologic and clinical subtypes of autoimmune pancreatitis: The honolulu consensus document. Pancreas. 2010;39:549-54.
25. Kamisawa T, Chari ST, Lerch MM, Kim MH, Gress TM, Shimosegawa T. Recent advances in autoimmune pancreatitis: Type 1 and type 2. Gut. 2013;62:1373-80.

Supplemental Table A1

Definite and probable chronic pancreatitis (CP) according to the M-ANNHEIM criteria
The diagnosis of chronic pancreatitis requires: I. a typical clinical history of chronic pancreatitis*: such as recurrent pancreatitis or abdominal pain II. One or more of the following additional criteria: * Except for primary painless pancreatitis.
For definite chronic pancreatitis: 1. Pancreatic calcifications. 2. Moderate or marked ductal lesions (according to the Cambridge classification). 3. Marked and persistent exocrine insufficiency defined as pancreatic steatorrhea markedly reduced by enzyme supplementation. 4. Typical histology of an adequate histological specimen.
For probable chronic pancreatitis: 1. Mild ductal alterations (according to the Cambridge classification) 2. Recurrent or persistent pseudocysts 3. Pathological test of pancreatic exocrine function (such as fecal elastase-1 test, secretin test, secretin–pancreozymin test) 4. Endocrine insufficiency (i.e., abnormal glucose tolerance test)

Supplemental Table A2

Diagnosis of chronic pancreatitis (CP) according to Büchler criteria
The diagnosis of chronic pancreatitis requires: – At least 1 clinical criteria – At least 1 additional criteria (imaging or pancreatic function)
Clinical criteria: – Abdominal pain – Recurrent attacks of pancreatitis – Steatorrhoe – Diabetes mellitus – One or more complications of CP: ▪ Bile duct obstruction/stenosis with cholestasis or jaundice ▪ Duodenal obstruction/stenosis with clinical signs ▪ Vascular obstruction/stenosis with clinical or morphological signs of portal/splenic vein hypertension ▪ Pancreatic pseudocysts with clinical signs (compression of adjacent organs, infection, bleeding, etc.) ▪ Pancreatic fistula (internal or external) ▪ Pancreatogenic ascites ▪ Other rare complications related to organs in vicinity (i.e., colonic stenosis, splenic pseudocyst)
Additional criteria: **Imaging findings** I. Ductal changes, defined as: – Irregularity of the main pancreatic duct or side branches, or – Calculi, or – Duct obstruction (strictures), or – Duct dilations (>3mm) II. Parenchymal changes, defined as: – General or focal enlargement of the gland, or – Cysts, or – Calcifications, or – Heterogeneous reflectivity **Direct pancreatic function test** – Abnormal secretin pancreozymin test or similar

Supplemental Table A3

Lüneburg criteria for diagnosis of chronic pancreatitis (CP)	
Parameter	**Score***
Morphological examinations	
– Post mortem diagnose of CP	4
– Histology	4
– Intra-operative findings, characteristics of CP	4
– Pancreatic calcifications, shown by any imaging procedure	4
Exocrine pancreatic function tests	
– Abnormal secretin pancreozymin test	3
– Abnormal pancreolauryl test	2
– Abnormal fecal chymotrypsin level	2
– Abnormal fecal elastase-1 level	2
– Steatorrhea	1
Imaging procedures	
– Abnormal ultrasound **	3
– Abnormal endoscopic ultrasound **	3
– Abnormal computed tomography **	3
– Abnormal ERCP **	3

ERCP, endoscopic retrograde cholangiopancreatography

* ≥ 4 Points, proven chronic pancreatitis. ** According to the Cambridge classification