

Diagnostic Accuracy of Transabdominal Ultrasound and Computed Tomography in Chronic Pancreatitis: A Head-to-Head Comparison



Authors

Ingrid Kvåle Nordaas^{1,2} , Trond Engjom^{1,2}, Odd Helge Gilja^{1,2}, Roald Flesland Havre^{1,2} , Dag André Sangnes^{1,2}, Ingfrid S. Haldorsen^{2,3}, Georg Dimcevski^{1,2}

Affiliations

- 1 National Centre for Ultrasound in Gastroenterology, Department of Medicine, Helse Bergen HF, Haukeland University Hospital, Bergen, Norway
- 2 Department of Clinical Medicine, University of Bergen, Bergen, Norway
- 3 Mohn Medical Imaging and Visualization Centre, Department of Radiology, Helse Bergen HF, Haukeland University Hospital, Bergen, Norway

Correspondence

Ingrid Kvåle Nordaas
Helse Bergen HF, National Centre for Ultrasound in Gastroenterology, Department of Medicine, Haukeland University Hospital
5021 Bergen
Norway
Tel.: +47 91588304, ingrid.kvale.nordaas@helse-bergen.no

Key words

chronic pancreatitis, CT, pancreas, pancreatitis, ultrasound

received 05.03.2021

revised 02.06.2021

accepted 19.06.2021

Bibliography

Ultrasound Int Open 2021; 7: E35–E44

DOI 10.1055/a-1542-9146

ISSN 2199-7152

© 2021. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany

ABSTRACT

Purpose Computed tomography (CT) is the most used imaging modality for diagnosing chronic pancreatitis (CP), but advances in transabdominal ultrasound (US) technology have given US a position as a viable alternative. We aimed to evaluate the diagnostic accuracy of abdominal CT and pancreatic US compared to the reference standard, a modified Mayo score.

Materials and Methods CT, US, and endoscopic ultrasound (EUS) were performed in patients referred due to suspected CP. The modified Mayo score included EUS results, clinical presentation, and results from exocrine and endocrine pancreatic function tests. We scored CT findings according to the modified Cambridge classification and US findings according to the Rosemont classification.

Results In total, 73 patients were included. 53 patients (73%) were categorized as CP and 20 (27%) as non-CP. CT and US yielded similar sensitivities (68% and 64%, respectively) and specificities (75 and 85%, respectively) and similar areas under the receiver operating characteristic curves for diagnosing CP. We found no significant differences between the areas under the receiver operating characteristic curves (AUROCs) for CT (AUROC 0.75, 95% CI 0.63–0.87) and US (AUROC 0.81, 95% CI 0.71–0.91).

Conclusion We conclude that CT and US had comparable, moderate accuracy in diagnosing CP. Neither modality had high enough sensitivity to exclude the diagnosis as a standalone method.

Introduction

Chronic pancreatitis (CP) is a fibro-inflammatory syndrome in which recurrent inflammatory episodes of the pancreas result in normal pancreatic parenchyma being replaced by fibrous connective tis-

sue [1–4]. Typical imaging features are calcifications, ductal irregularities and dilatations, increased parenchymal density or echogenicity, gland atrophy, and pseudocysts [1]. Abdominal pain is the most frequently reported symptom of CP, and pain patterns vary

from intermittent pain episodes to more prominent, continuous pain syndromes [1]. CP is further characterized by exocrine and endocrine pancreatic insufficiency, contributing to bloating, steatorrhea, underweight, malnutrition, and metabolic bone disease [1, 4, 5].

CP can be particularly difficult to diagnose in the early phases, when patients may have recurring inflammatory episodes, either silent or accompanied by non-specific symptoms, but classical structural changes or functional pancreatic insufficiencies are not yet detectable. Diagnostic guidelines [1, 2, 4–6] recommend a multimodal workup, including diagnostic imaging, pancreatic function testing and symptom evaluation, and several diagnostic criteria have been developed [7–11]. Computed tomography (CT), endoscopic ultrasound (EUS) and magnetic resonance imaging (MRI) are all recommended imaging modalities as part of the diagnostic workup for CP [1, 5, 6]. Transabdominal ultrasound (US) is widely used and recommended as a first-line modality in subjects with abdominal pain [1, 2, 12]. Although opinions on the role of US in CP diagnostics are divided, US is recommended in patients with advanced CP and when repeated examinations are warranted [1, 2, 5, 12]. The imaging modalities used to diagnose CP have inherent weaknesses related to operator and patient dependencies, and some modalities involve invasive procedures (EUS) or ionizing radiation (CT) or have limited ability to detect key features such as calcifications (MRI) or early ductal changes (CT and US) [6, 12, 13]. By combining different modalities and applying them in the appropriate clinical settings, a multimodal imaging workup may reduce the impact of modality limitations and improve diagnostic accuracy.

Image scoring systems may give more weight to important or pathognomonic features, e. g., calcifications in CP [1] than to less important features. Thus, a well-designed weighted scoring system should ideally increase the diagnostic accuracy for diagnosing CP compared to just counting the number of positive features. The Cambridge classification was originally used for endoscopic retrograde pancreatography, and later adapted to MRI and CT [14, 15]. EUS and US have often been scored by counting the number of features without any weighting [16, 17], but in recent years, the Rosemont classification [16] has proven to be a promising weighted scoring system for these modalities.

CT is the first-line imaging modality when CP is suspected, but the advances in ultrasound technology and image quality [18, 19] have strengthened the potential role of US as a first-line imaging modality. To the authors' knowledge, no studies have evaluated the diagnostic performance of CT and US for diagnosing CP against a common reference standard since the 1990s. Thus, in this head-to-head analysis in patients with suspected CP, we aimed to compare the diagnostic accuracy of modern CT and US in comparison with a multimodal diagnostic score for CP [7] based on findings from EUS, clinical presentation and results from exocrine and endocrine pancreatic function tests. We applied both unweighted scores and the most frequently used weighted scoring systems: the modified Cambridge classification for CT and the Rosemont classification for EUS and US

► **Table 1–6.**

Materials and Methods

Study design

The Bergen Pancreas Database collected information on 141 patients referred to our outpatient clinic with symptoms or imaging findings suggestive of CP. Patients were included consecutively in the period from 2009 to 2016. Patient characteristics, results from biochemistry and pancreatic function tests, and results from various imaging examinations were recorded. In this paper we present analyses of the diagnostic imaging data collected upon inclusion, applying a cross-sectional design. We included all patients who had a complete imaging workup, including CT, US, and EUS. Patients were excluded if image quality/visualization was insufficient in any of the three examinations.

Diagnostic standards

The reference standard was a modified version of the Mayo score [7] based on EUS, clinical presentation, and exocrine and endocrine function status (► **Table 1**). To avoid the diagnostic standard being impacted by the tested modalities, CT and US were not included in the score. The EUS imaging features that were used were parenchymal calcifications, ductal stones, dilated or irregular main pancreatic duct contour, dilated side ducts, pseudocysts, and honeycomb-patterned lobulation. Four points were given if parenchymal calcifications or ductal stones were present, and 3 points were given if any of the other features were present. Patients with Mayo score ≥ 4 were categorized as CP, and patients with scores < 4 were categorized as non-CP. EUS is the recommended modality for diagnosing early CP. However, our reference standard did not include an evaluation of early CP.

EUS imaging

EUS examinations were performed by an experienced operator (R.F.H.) using a linear EG-3870 UTK or radial EG-3670 URK scope (Pentax Medical, Pentax Europe, Hamburg, Germany), with frequencies 7.5–12 MHz. We used a Hitachi Ascendus scanner (Hitachi Medical Systems Europe, Zug, Switzerland). Patients were offered intravenous conscious sedation as per local guidelines (Midazolam and Pethidine or Fentanyl). The examination was performed by scanning the pancreatic body and tail from the gastric ventricle, and then scanning the pancreatic head under slow retraction from the second part of the duodenum. The findings were registered according to definitions from the Rosemont classification [16] (► **Table 1**). In addition to being part of the reference standard, findings from EUS were included in the single feature agreement analyses to allow for the comparison of CT versus US versus EUS. Scoring of EUS was performed immediately after examinations, and the operator was blinded to the patients' medical history, other imaging, and test results.

US imaging and scoring

Transabdominal US examinations were performed by a gastroenterologist (T.E.) with > 10 years of experience in pancreatic ultrasound and imaging. Examinations and scoring were performed with blinding to the patients' medical history, other imaging, and test results. Scoring was performed immediately after examinations. US was carried out using a GE Logic E9 scanner (GE Healthcare, Chicago, IL, USA) with a 1–5 MHz curvilinear probe, and when possible,

► **Table 1** Diagnostic scoring systems for chronic pancreatitis

The modified Mayo score	
The CP diagnosis requires ≥ 4 points	
4 points	Pancreatic calcifications on EUS or typical histologic findings
3 points	Moderate or marked morphologic changes on EUS *
2 points	Reduced exocrine pancreatic function †
2 points	History of acute pancreatitis or upper abdominal pain
1 point	Diabetes mellitus, glycated hemoglobin ≥ 48 mmol/mol.
The modified Cambridge classification for CT	
Cambridge grade	CT features
0 - Normal	None
1 - Equivocal	Cannot be delimited on CT with current methods
2 - Mild	2 or more of the following changes: – Pancreatic duct between 2 and 4 mm in the body of the pancreas ‡ – Heterogeneous parenchymal structure – Small cystic changes (< 10 mm) – Duct irregularities (1-2 mm variation) – 3 or more pathological side ducts (Excluded criteria: slight enlargement of the pancreas)
3 - Moderate	All changes specified in 2 plus pathological main duct (>4 mm)
4 - Marked	One of the changes specified in 2 and 3 plus one or more of the following: – Cystic structures (≥ 10 mm) – Parenchymal calcifications – Intraductal filling defects (chalk stones) – Duct obstruction (strictures) – Severe duct irregularities (>2 mm variation)
The Rosemont classification for US	
Parenchymal features	
Hyperechoic foci with shadowing	Major A
Lobularity	
A: without honeycombing	Minor
B: with honeycombing	Major B
Hyperechoic foci without shadowing	Minor
Cysts	Minor
Stranding	Minor
Ductal features	
MPD calculi	Major A
Irregular MPD contour	Minor
Dilated side branches	Minor
MPD dilatation	Minor
Hyperechoic MPD margin	Minor
1 - Consistent with CP	A. 1 major A feature + ≥ 3 minor features B. 1 major A feature + 1 major B feature C. 2 major A features
2 - Suggestive of CP	A. 1 major A feature + <3 minor features B. 1 major B feature + ≥ 3 minor features C. ≥ 5 minor features (any)
3 - Indeterminate for CP	A. 3 to 4 minor features, no major features B. Major B feature alone or with <3 minor features
4 - Normal	≤ 2 minor features, no major features
<p>Mayo score [7] was modified not to incorporate findings from CT and US, and it was used as the reference standard for the CP diagnosis. We used the modified Cambridge classification [15] to score results from CT imaging and the Rosemont classification [16] for US. CP= chronic pancreatitis; CT= computed tomography; EUS= endoscopic ultrasonography; MPD= main pancreatic duct; US= transabdominal ultrasonography. * Moderate or marked morphologic changes on EUS were defined as the presence of one or more of the following features: irregular or dilated main pancreatic duct contour, dilated side ducts, pseudocysts and honeycomb-patterned lobulation. † Reduced exocrine pancreatic function by endoscopic short test or fecal elastase 1 level. Cut-offs: fecal elastase 1 <200 $\mu\text{g/g}$ and endoscopic short test, peak value of bicarbonate concentration <80 mmol/L. ‡ The typical definition of duct dilatation in the body of the pancreas uses a cut-off ≥ 3 mm. Patients scored positive if the duct diameter was ≥ 3 mm and ≤ 4 mm.</p>	

a 9 MHz linear probe. Patients were examined after overnight fasting to optimize visualization. US was performed with patients in a supine or right lateral position, and the probe were placed in transverse and oblique positions in the lateral/posterior left subcostal region. Standardized abdominal US settings were applied: Frequency 4.0 MHz (curvilinear probe) and 9.0 MHz (linear probe), dynamic range 34, and frame rate 15–22 frames per second. US examinations were completed in B-mode and supplemented with color Doppler to evaluate, *e. g.*, twinkling artifacts around edged calcifications. Scanning depth varied depending on the patient anatomy. US visualization of the pancreatic head, body, and tail was graded from 1 to 4 (1 = good, 2 = adequate, 3 = poor, and 4 = not visible), and patients were excluded if the visualization was 3 or higher in all segments.

The Rosemont classification (► **Table 1**) originally developed for EUS was used to score ductal and parenchymal features on US [16, 19]. The Rosemont categories for US are: 1 = consistent with CP, 2 = suggestive of CP, 3 = indeterminate for CP, and 4 = normal. In the Rosemont classification, the cut-off for the CP diagnosis is ≤ 2 . We also counted the number of positive single criteria from the Rosemont classification (unweighted US score), using a cut-off ≥ 2 according to a previous publication [19]. ► **Fig. 1** shows a US image in a patient with CP.

CT imaging and scoring

CT scans were performed using standard abdominal or pancreatic CT protocols on clinical hospital scanners. Intravenous contrast was administered in 97% (71/73) of the CT scans. Abdominal protocols included scanning in the portal venous phase, and pancreatic protocols included scans pre-contrast and in the late arterial and portal venous phase. Scans were stored and re-evaluated for the purpose of this study. Image quality was evaluated by two observers (T.E., gastroenterologist, and I.K.N., medical doctor), and patients were excluded if the quality was deemed insufficient. CT imaging was scored by I.K.N., who was blinded to the patients' medical history, other imaging, and test results. Parenchymal and ductal features were assessed and scored using the modified Cambridge classification (► **Table 1**) [15], following predefined reading standards. The classification categorizes patients into five groups (0 = normal,

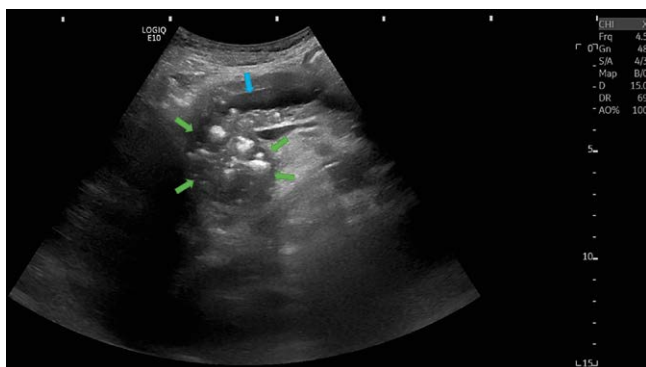
1 = equivocal, 2 = mild, 3 = moderate, and 4 = marked). We also applied an unweighted CT score, where the score was equal to the number of positive single criteria from the Cambridge classification. The cut-off was ≥ 2 according to a previous publication [20]. ► **Figure 2** shows a CT image in a patient with CP.

Statistical analysis

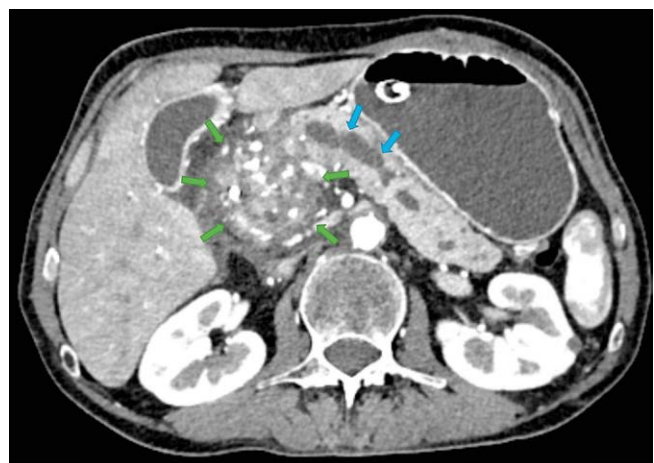
Statistical analyses were performed using IBM SPSS Statistics, version 26.0 (IBM, Armonk, NY, USA). Data are presented as mean \pm standard deviation (SD) or as median with interquartile range (IQR). Normality distributions were tested using Q-Q plots and Shapiro-Wilk's test. When comparing groups, we used independent samples t-test and Mann-Whitney U test for parametric and non-parametric continuous data, respectively. Pearson's χ^2 -test was used for binary and categorical data. We applied Fisher's Exact Probability test if the minimum expected cell count was < 5 in any cells when comparing frequencies between groups. Results on diagnostic accuracy are presented as area under the receiver operating characteristic curves (AUROCs), sensitivity, and specificity with 95% confidence intervals. Differences in AUROCs were evaluated using a paired-sample AUROCs comparison in SPSS. Cohen's kappa (κ) was used to evaluate the agreement for binary variables, and weighted kappa was used for ordinal variables. The level of agreement was defined as follows: poor ($\kappa < 0.20$), fair ($\kappa = 0.21$ – 0.40), moderate ($\kappa = 0.41$ – 0.60), good ($\kappa = 0.61$ – 0.80), and very good ($\kappa = 0.81$ – 1). A value of $p < 0.05$ was considered statistically significant.

Ethics

The study was conducted according to the Helsinki Declarations and approved by the Regional Committees for Medical and Health Research Ethics, Western Norway (REK-Vest, registration numbers 2011/590 and 2019/1037). All patients received written and oral information about the study and signed an informed consent form before inclusion in the database and prior to any study related examinations. This paper adheres to the Standards for Reporting of Diagnostic Accuracy (STARD) [21].



► **Fig. 1** US image illustrating a large inflammatory mass (green arrows) with calcifications, and an irregular and dilated main pancreatic duct (blue arrow) in a patient with chronic pancreatitis. Fig. 2 shows a CT image of the same patient.



► **Fig. 2** CT image illustrating a large inflammatory mass (green arrows) with calcifications, and an irregular and dilated main pancreatic duct (blue arrows) in a patient with chronic pancreatitis. Fig. 1 shows a US image of the same patient.

Results

Participants

From the database cohort comprised of 141 patients, 84 patients had completed both CT, EUS and US, and were considered eligible for the study. We excluded nine patients due to insufficient visualization on US and two patients due to poor CT image quality. In total, 73 patients were included in the analyses. 53 patients (73 %) had a Mayo score ≥ 4 , fulfilling the diagnostic criteria of CP. The remaining 20 patients (27 %) were categorized as non-CP (► Fig. 3). Non-CP patients were diagnosed with other conditions explaining the symptoms, including recurring acute pancreatitis, bile stone disease, and functional dyspepsia. The CP group had a significantly lower body mass index and fecal elastase levels than the non-CP group, but there were no significant differences in age, gender, or glycated hemoglobin (► Table 2). Pancreatic CT protocols were used in 47 % (34/73) and abdominal CT protocols in 53 % (39/73). The use of CT protocols (abdominal vs. pancreatic) was not significantly different between CP and non-CP patients ($p = 0.223$).

Visualization

On US, the entire pancreas was visualized in 62 % of the included patients. The pancreatic tail was sufficiently visualized (scores 1 or 2) in 65 %, the body in 97 %, and the head in 89 %.

Diagnostic performance indices

Diagnostic performance indices for CT and US were calculated using the modified Mayo score as the reference standard: CT (Cambridge classification, cut-off ≥ 2) yielded a sensitivity of 68 % and a specificity of 75 %, and US (Rosemont classification, cut-off ≤ 2) yielded a sensitivity of 64 % and a specificity of 85 % (► Table 3).

For CT, the areas under the receiver operating characteristic curves (AUROCs) were lower when using the Cambridge classifica-

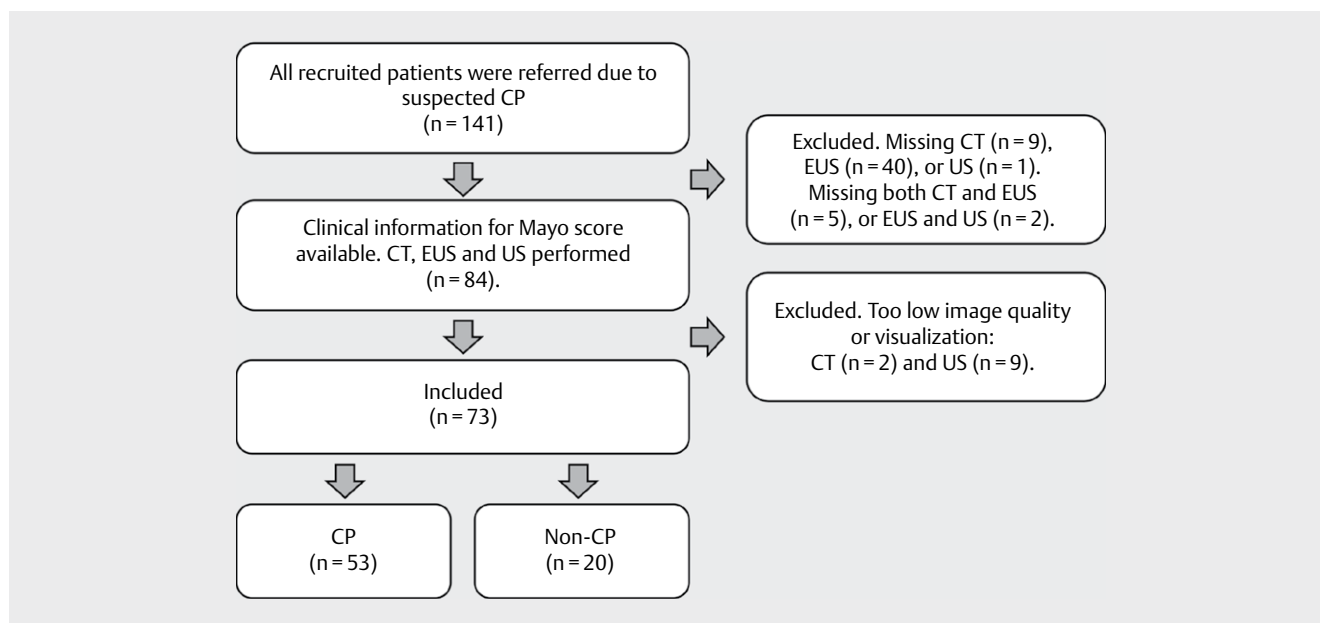
tion (0.75, 95 % CI 0.63–0.87) compared to when using the unweighted CT score (0.80, 95 % CI 0.70–0.90), $p = 0.05$. The AUROCs for US were 0.81 (95 % CI 0.71–0.91) for the Rosemont classification and 0.84 (95 % CI 0.74–0.94) for the unweighted US score, $p = 0.188$. Any differences in AUROCs between CT and US were non-significant ($p > 0.05$). ROC curves are presented in ► Fig. 4.

Agreement analyses

We found moderate agreement ($\kappa = 0.45$) for the CP diagnosis between CT and US when using the modified Cambridge and the Rosemont classifications, and fair agreement ($\kappa = 0.37$) when using the unweighted scores. We found moderate agreement (weighted $\kappa = 0.43$) between the categories from the modified Cambridge classification and the inverted Rosemont classification. The most frequent feature was calcifications, which were present in 51 % of CP patients on CT, 64 % on US and 59 % on EUS. We found good agreement for calcifications between CT and US ($\kappa = 0.75$) and between CT and EUS ($\kappa = 0.69$), and very good agreement between US and EUS ($\kappa = 0.84$). All agreement analyses are presented in ► Table 4. Frequencies of CP-related imaging features on CT and US are presented in ► Table 5 and ► Table 6, respectively.

Discussion

In this head-to-head study, we aimed to compare the diagnostic accuracy of CT and US in patients with suspected CP. The reference standard was a modified Mayo score based on findings from EUS, clinical presentation and results from exocrine and endocrine pancreatic function tests. Our results showed that CT and US yielded comparable moderate diagnostic accuracies for diagnosing CP and a moderate agreement between the imaging scoring systems: CT had a sensitivity of 68 % and specificity of 75 %, and US had a sensitivity of 64 % and a specificity of 85 %. The agreement for calcifica-



► Fig. 3 Flowchart showing patient enrollment. CP = chronic pancreatitis; CT = computed tomography; EUS = endoscopic ultrasound; US = ultrasound.

► **Table 2** Patient characteristics and imaging scoring results in a multimodal imaging study on chronic pancreatitis

	Non-CP (n = 20)	CP (n = 53)	Missing (n)	p
Females, n (%)	10 (50)	33 (62)	0	ns
Age, mean (SD)	54 (13)	54 (13)	0	ns
Fecal elastase 1 (µg/g), median (IQR)	433 (270–500)	159 (39–500)	6	0.001
PEI frequency, n (%)	2 (10)	27 (57)	6	<0.001
BMI (kg/m ²), median (IQR)	24.3 (19.9–27.5)	21.2 (18.4–24.1)	8	0.040
Underweight frequency, n (%)	0	12 (25)	8	0.022
HbA1c (mmol/mol), median (IQR)	5,5 (5,2–5,9)	5,7 (5,4–6,5)	3	ns
Diabetes frequency, n (%)	2 (11)	12 (23)	3	ns
Smokers frequency, n (%)	3 (17)	26 (51)	4	0.011
Alcohol consumption, frequencies, n (%)				
Current non-drinkers	8 (47)	27 (55)	7	ns
Current regular drinkers, ≥ 7 std. drinks per week	1 (6)	4 (8)	7	ns
CT Cambridge classification, n (%)			0	0.005
0 – normal	9 (45)	8 (15)		
1 – equivocal	6 (30)	9 (17)		
2 – mild	1 (5)	1 (2)		
3 – moderate	0	0		
4 – marked	4 (20)	35 (66)		
US Rosemont classification, n (%)			0	<0.001
1 – consistent with CP	1 (5)	22 (42)		
2 – suggestive of CP	2 (10)	12 (23)		
3 – indeterminate for CP	0	6 (11)		
4 – normal	17 (85)	13 (25)		
Unweighted CT score, median (IQR)	0 (0–2)	3 (1–5)	0	<0.001
Unweighted US score, median (IQR)	0 (0–1)	3 (1–5)	0	<0.001

Pancreatic exocrine insufficiency was defined as fecal elastase 1 <200 µg/g, underweight as BMI ≤ 18.5 kg/m² and diabetes as HbA1c ≥ 48 mmol/mol. Overall *p*-values from χ^2 -test are given for the Cambridge classifications (0–4) and the Rosemont classifications (1–4). BMI = body mass index; CP = chronic pancreatitis; CT = computed tomography; HbA1c = glycated hemoglobin; IQR = interquartile range; PEI = pancreatic exocrine insufficiency; SD = standard deviation; US = ultrasound.

tions was very good between US and EUS, and agreements between modalities for duct dilatations ranged from poor to moderate.

Previous studies on the diagnostic accuracy of CT and/or US in CP have shown divergent results, with sensitivities ranging from 58–100% for CT and 38–100% for US, and specificities ranging from 59–100% for CT and 34–100% for US [19, 20, 22–25]. The most recent study [19] reported good diagnostic accuracy of modern transabdominal US in CP, with a sensitivity of 81% and a specificity of 97% for the Rosemont classification, and a sensitivity of 69% and a specificity of 97% for the unweighted US score. The US scoring tools were identical, but the reference standard also included results from CT. An EUS-based reference standard has the potential to detect more cases of early CP. Because patients with early CP typically do not present with irreversible morphologic changes, increasing imaging sensitivity is associated with decreasing specific-

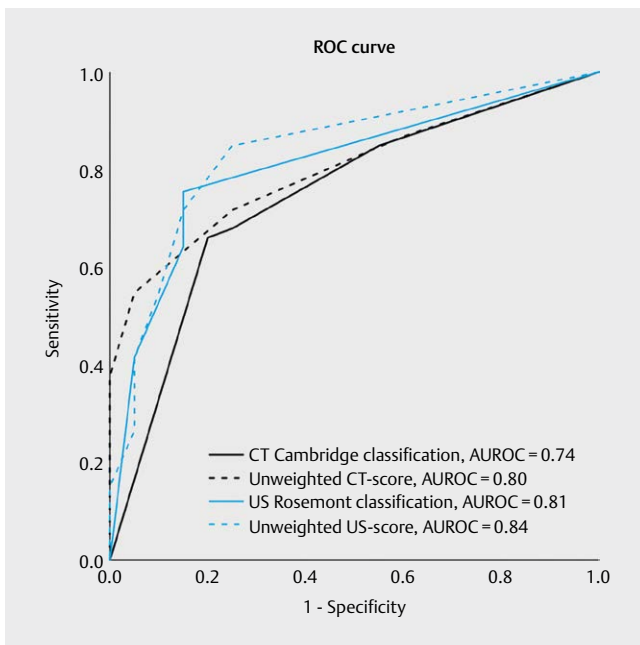
ity [5], possibly explaining why we now present lower diagnostic performance indices for US.

Issa et al. [22] published a systematic review on the diagnostic performance of different imaging modalities in CP. They presented pooled sensitivity and specificity data for CT, US, EUS, MRI, and endoscopic retrograde pancreatography, plus a head-to-head analysis of subjects who had undergone both CT and US. Their analyses showed that CT had a sensitivity of 75% and a specificity of 91%, and US had a sensitivity of 67% and a specificity of 98%. Their head-to-head analyses were particularly similar to our results. The majority of the studies on CT and/or US were published between 1977 and 1999 [22]. This complicates the comparison to our data for two reasons. First, CT and US technologies have evolved considerably during this period [22, 26], and second, with EUS facilitating the diagnosis of early CP [2], the diagnostic standards and disease severity among newly diagnosed patients have changed. All studies

► **Table 3** Diagnostic performance indices for diagnosing chronic pancreatitis

	AUROC	Cut-off	Sensitivity (%)	Specificity (%)
CT Cambridge	0.75 (0.63–0.87)	≥2	68 (54–80)	75 (51–91)
CT Unweighted	0.80 (0.70–0.90)	≥2	72 (58–83)	75 (51–91)
US Rosemont	0.81 (0.71–0.91)	≤2	64 (50–77)	85 (62–97)
US Unweighted	0.84 (0.74–0.94)	≥2	72 (58–83)	85 (62–97)

The table shows areas under the receiver operating characteristic curves (AUROCs) and diagnostic performance indices with 95% confidence intervals. AUROC for the Cambridge classification was significantly lower than AUROC for the unweighted CT score ($p = 0.050$). CT = computed tomography; US = ultrasound.



► **Fig. 4** ROC curve with areas under the ROC curves (AUROCs) for CT and US in diagnosing chronic pancreatitis. AUROC for the Cambridge classification was lower than AUROC for the unweighted CT score ($p = 0.050$). Other differences in AUROCs were non-significant (Cambridge vs. Rosemont: $p = 0.377$, Cambridge vs. unweighted US-score: $p = 0.173$, Rosemont vs. unweighted US-score: $p = 0.188$, Rosemont vs. unweighted CT-score: $p = 0.886$, unweighted US-score vs. unweighted CT-score: $p = 0.519$). CP = chronic pancreatitis; CT = computed tomography; US = ultrasound.

in the review recruited patients with suspected CP or suspected pancreatic disease, but the reference standard was only similar to ours in one of them [23]. There were also differences in scoring methods used for CT and US, and in some publications, the scoring methods were not available.

In the agreement analyses, we found that US and EUS had higher agreement on duct dilatations compared to CT and EUS, indicating that US may be better than CT in detecting ductal changes. However, the difference may also be explained by the slight difference in the definitions of duct dilatations between the Cambridge and Rosemont classifications. Calcifications are not as apparent on US as on CT, but our results confirm the ability of US to detecting calcifications: US has good and very good agreement with CT and EUS, respectively.

Though good results have been achieved with Rosemont for transabdominal US previously [19, 27], the score is complex, and a simpler US scoring tool, as recently proposed by Pagliari et al. [28], is warranted. The Cambridge classification for CT has several weaknesses, including its somewhat complex structure and large focus on ductal changes and other subtle changes. Given the current scoring systems, neither CT nor US has sufficient diagnostic accuracy to recommend one over the other.

Several factors influence the choice of imaging modality in the diagnostic workup of CP. CT and US are widely available and at relatively low costs. However, their other strengths and limitations differ. CT image quality was sufficient in 82 of the 84 patients we evaluated, and because CT performs well in visualizing the whole pancreas, it has an advantage in ruling out differential diagnoses and can function as a baseline examination [6, 12]. Due to ionizing radiation and risk associated with intravenous contrast agents, repeated use and use of CT in young patients or patients with impaired kidney function should be limited. US does not have such limitations and can easily be repeated if needed for frequent follow-up. Furthermore, bedside US enables the clinician to get immediate answers. Intestinal gas and obesity may, however, hinder adequate visualization on US. Importantly, the whole pancreas was only visualized in 62% of our patients, indicating that US should not be the sole modality if pancreatic cancer is suspected. Ruling out concomitant cancer in a pancreas structurally altered by CP is difficult [1, 2], and even though EUS may be the best modality to detect malignancies [1], a combination of different imaging modalities may be recommended, particularly when indicated based on clinical suspicion or known risk factors [2]. The full potential for advanced US in diagnosing CP is still unclear, but contrast-enhanced US reportedly improves the diagnostic accuracy when characterizing focal pancreatic lesions [1, 29], and US elastography may also prove beneficial [30].

Our findings provide a reminder of the differences in strengths and limitations connected to each modality, pinpointing the importance of a multimodal and individually adapted approach. Further exploration of strengths and limitations of CT and US is warranted, and combined with the development and validation of updated scoring systems, this may provide better diagnostics for patients with CP.

Limitations

The conclusions regarding diagnostic accuracy are highly dependent on the scoring systems used to evaluate diagnosis and imaging, and the results are limited by the strengths and weaknesses in these mainly expert opinion-based systems. None of the systems are properly validated, and inaccuracies in any one of the three scoring systems will have implications for results and further inter-

► **Table 4** Agreement analyses on diagnosis and imaging parameters in patients with chronic pancreatitis

				Agreement		
				%	κ	
CP diagnosis	CT Cambridge classification	vs.	US Rosemont classification	77	0.51	Moderate
	CT Cambridge classification	vs.	Unweighted CT score	97	0.94	Very good
	US Rosemont classification	vs.	Unweighted US score	81	0.62	Good
	Unweighted CT score	vs.	Unweighted US score	75	0.50	Moderate
Calcifications	CT	vs.	US	84	0.67	Good
	CT	vs.	EUS	84	0.66	Good
	US	vs.	EUS	92	0.84	Very good
Pseudocysts	CT	vs.	US	81	0.35	Fair
	CT	vs.	EUS	74	0.36	Fair
	US	vs.	EUS	72	0.32	Fair
Duct dilatations	CT	vs.	US	64	0.30	Fair
	CT	vs.	EUS	63	0.28	Fair
	US	vs.	EUS	75	0.48	Moderate
				Weighted agreement		
Scores	CT Cambridge classification	vs.	US Rosemont classification		0.43	Moderate

Agreement between scoring systems and modalities for diagnosis and three key features. Results are presented as percent agreement (%) and Cohen's kappa (κ). Calcifications include both parenchymal and ductal calcifications. CP = chronic pancreatitis; CT = computed tomography; EUS = endoscopic ultrasound; US = transabdominal ultrasound.

► **Table 5** CP-related imaging features on CT in a cohort of patients with suspected chronic pancreatitis

	Total	Non-CP	CP	Missing	<i>p</i>
MPD dilatation 2–4 mm, pancreatic body	26 (37)	6 (32)	20 (39)	2	ns
Heterogeneous parenchymal structure	38 (52)	5 (25)	33 (62)	0	0.004
Small cystic changes	5 (7)	2 (10)	3 (6)	0	ns
Duct irregularities	12 (17)	1 (5)	11 (21)	2	ns
Side ducts	0	-	-	0	-
MPD >4 mm	15 (21)	0	15 (29)	1	0.007
Large cystic structures	14 (19)	2 (10)	12 (23)	0	ns
Parenchymal calcifications	28 (38)	2 (10)	26 (49)	0	0.003
Intraductal filling defects	19 (26)	0	19 (36)	0	0.001
Duct obstructions	16 (22)	0	16 (31)	1	0.004
Severe duct irregularities	13 (18)	0	13 (25)	1	0.014
All calcifications	29 (40)	2 (10)	27 (51)	0	0.001

The frequencies of positive single features from CT examinations are presented as number of cases (%). Missing data are expressed as number of cases. Definitions according to the modified Cambridge classification for CT. CP = chronic pancreatitis; CT = computed tomography; MPD = main pancreatic duct.

► **Table 6** CP-related imaging features on transabdominal US in a cohort of patients with suspected chronic pancreatitis

	Total	Non-CP	CP	Missing	p-value
Hyperechoic foci with shadowing	36 (49)	3 (15)	33 (62)	0	<0.001
Lobularity (all)	2 (3)	0	2 (4)	0	ns
Hyperechoic foci without shadowing	13 (18)	2 (10)	11 (21)	0	ns
Cysts	10 (14)	1 (5)	9 (17)	0	ns
Stranding	19 (26)	1 (5)	18 (34)	0	0.015
MPD calculi	17 (23)	0	17 (32)	0	0.004
Irregular MPD contour	32 (44)	3 (15)	29 (55)	0	0.003
Dilated side branches	18 (25)	1 (5)	17 (32)	0	0.016
MPD dilatation	22 (30)	0	22 (42)	0	<0.001
Hyperechoic MPD margin	4 (5)	1 (5)	3 (6)	0	ns
Calcifications (all)	37 (51)	3 (15)	34 (64)	0	<0.001

The frequencies of positive single features from US examinations are presented as number of cases (%). Definitions according to the Rosemont classification. CP = chronic pancreatitis; MPD = main pancreatic duct; US = ultrasound.

pretation. We applied the most frequently used systems and included unweighted scores to offer an alternative to the weighted systems.

Because of the similarities between EUS and US with regards to technology and scoring systems, using EUS as part of the reference standard may have given US an advantage compared to CT. Histology from, e. g., fine needle aspiration or fine needle biopsies would provide a better reference standard. However, due to the risk of complications related to such procedures, this could not be justified in our study population.

The study protocol was based on a consecutive intention-to-diagnose design, aiming to complete all the required imaging (CT, EUS, and US). Still, for various reasons, some patients failed to complete all imaging examinations. This may have created selection biases connected to individual patient characteristics or disease presentation. The results for US do not take into account the cases (n = 9) in which the pancreas could not be sufficiently visualized, and exclusion caused by bowel gas or obesity may also have caused a selection bias.

Due to the difference in experience levels between the observers scoring US and CT, CT examinations were interpreted by both observers. Analyses showed good interobserver agreement for scoring ($\kappa = 0.66$) and diagnosis ($\kappa = 0.67$) and no significant differences in AUROCs, indicating that this minimally affected the CT scoring results. Visualization assessments for EUS and US examinations were performed by single observers, while the quality of CT examinations were assessed by two observers. Both CT and US examinations were performed blinded to patient history and other examinations, but blinding for the general reason for referral (CT/US) plus patient appearance (US) was not feasible and may have produced blinding bias.

Operator experience level is relevant in pancreatic US, and diagnostic performance cannot be directly translated to a setting with an inexperienced operator. All imaging modalities require training

to achieve a certain level of reliability. Particularly EUS has a long learning curve and can only be performed by skilled personnel [19].

Conclusion

In this head-to-head study, CT and US yielded similar, but only moderate diagnostic performance indices, not high enough to support that they should be used as single modalities. CT plays an important role in evaluating complications and differentiating CP from other diagnoses, and it is still the modality of choice in the initial diagnostic workup of CP. US does not have limitations related to ionizing radiation and CT contrast agents, and this study shows that US has comparable accuracy to that of CT given sufficient visualization. A combination of both CT and US may be beneficial in the primary workup for CP, and US seems particularly favorable for repeated examinations and follow-up in CP. The advantages and drawbacks of each modality are different, and the modality of choice should match the requirements in each patient's case.

Acknowledgements

We would like to thank the gastrointestinal unit nurses for their assistance during EUS examinations, and our laboratory technicians Liv Aasmul and Ingeborg Brønstad who analyzed duodenal juice for pancreatic exocrine function testing. I.K.N. and D.A.S. have received PhD grants from the Western Norway Regional Health Authority.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

- [1] Lohr JM, Dominguez-Munoz E, Rosendahl J, Besselink M, Mayerle J, Lerch MM et al. United European Gastroenterology evidence-based guidelines for the diagnosis and therapy of chronic pancreatitis (HaPanEU). *United European Gastroenterol J*. 2017; 5: 153–199
- [2] Hoffmeister A, Mayerle J, Beglinger C, Buchler MW, Bufler P, Dathe K et al. English language version of the S3-consensus guidelines on chronic pancreatitis: Definition, aetiology, diagnostic examinations, medical, endoscopic and surgical management of chronic pancreatitis. *Z Gastroenterol*. 2015; 53: 1447–1495
- [3] Kleeff J, Whitcomb DC, Shimosegawa T, Eposito I, Lerch MM, Gress T et al. Chronic pancreatitis. *Nat Rev Dis Primers*. 2017; 3: 17060
- [4] Beyer G, Habtezion A, Werner J, Lerch MM, Mayerle J. Chronic pancreatitis. *Lancet*. 2020; 396: 499–512
- [5] Gardner TB, Adler DG, Forsmark CE, Sauer BG, Taylor JR, Whitcomb DC. ACG Clinical Guideline: Chronic Pancreatitis. *Am J Gastroenterol*. 2020; 115: 322–339
- [6] Frokjaer JB, Akisik F, Farooq A, Akpinar B, Dasyam A, Drewes AM et al. Guidelines for the Diagnostic Cross Sectional Imaging and Severity Scoring of Chronic Pancreatitis. *Pancreatol*. 2018; 18: 764–773
- [7] Layer P, Yamamoto H, Kalthoff L, Clain JE, Bakken LJ, DiMaggio EP. The different courses of early- and late-onset idiopathic and alcoholic chronic pancreatitis. *Gastroenterology*. 1994; 107: 1481–1487
- [8] 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. *J Gastroenterol*. 2007; 42: 101–119
- [9] Buchler MW, Martignoni ME, Friess H, Malfertheiner P. A proposal for a new clinical classification of chronic pancreatitis. *BMC Gastroenterol*. 2009; 9: 93
- [10] Bagul ASA. Evaluation of the Manchester classification system for chronic pancreatitis. *JOP*. 2006; 390–396
- [11] Lankisch PG, Assmus C, Maisonneuve P, Lowenfels AB. Epidemiology of pancreatic diseases in Luneburg County. A study in a defined german population. *Pancreatol*. 2002; 2: 469–477
- [12] Conwell DL, Lee LS, Yadav D, Longnecker DS, Miller FH, Morteale KJ et al. American Pancreatic Association Practice Guidelines in Chronic Pancreatitis: evidence-based report on diagnostic guidelines. *Pancreas*. 2014; 43: 1143–1162
- [13] Dasyam AK, Vipperla K, Slivka A, Gong T, Papachristou GI, Whitcomb DC et al. Computed tomography based scoring system in a prospectively ascertained cohort of patients with chronic pancreatitis. *Pancreatol*. 2019; 19: 1027–1033
- [14] Sarner M, Cotton PB. Classification of pancreatitis. *Gut*. 1984; 25: 756–759
- [15] Schreyer AG, Jung M, Riemann JF, Niessen C, Pregler B, Grenacher L et al. S3 guideline for chronic pancreatitis-diagnosis, classification and therapy for the radiologist. *Rofo*. 2014; 186: 1002–1008
- [16] Catalano MF, Sahai A, Levy M, Romagnuolo J, Wiersema M, Brugge W et al. EUS-based criteria for the diagnosis of chronic pancreatitis: the Rosemont classification. *Gastrointest Endosc* 2009; 69: 1251–1261
- [17] Stevens T, Lopez R, Adler DG, Al-Haddad MA, Conway J, Dewitt JM et al. Multicenter comparison of the interobserver agreement of standard EUS scoring and Rosemont classification scoring for diagnosis of chronic pancreatitis. *Gastrointest Endosc*. 2010; 71: 519–526
- [18] Whitsett M. Ultrasound Imaging and Advances in System Features. *Ultrasound Clinics*. 2009; 4: 391–401
- [19] Engjom T, Sangnes DA, Havre RF, Erchinger F, Pham KD, Haldorsen IS et al. Diagnostic Accuracy of Transabdominal Ultrasound in Chronic Pancreatitis. *Ultrasound Med Biol*. 2017; 43: 735–743
- [20] Nordaas IK, Dimcevski G, Gilja OH, Havre RF, Haldorsen IS, Engjom T. Diagnostic accuracy of computed tomography scores in chronic pancreatitis. *Pancreas*. 2021: (in press)
- [21] Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016; 6: e012799
- [22] Issa Y, Kempeneers MA, van Santvoort HC, Bollen TL, Bipat S, Boermeester MA. Diagnostic performance of imaging modalities in chronic pancreatitis: a systematic review and meta-analysis. *Eur Radiol*. 2017; 27: 3820–3844
- [23] Buscaill L, Escourrou J, Moreau J, Delvaux M, Louvel D, Lapeyre F et al. Endoscopic ultrasonography in chronic pancreatitis: a comparative prospective study with conventional ultrasonography, computed tomography, and ERCP. *Pancreas*. 1995; 10: 251–257
- [24] Swobodnik W, Meyer W, Brecht-Kraus D, Wechsler JG, Geiger S, Malfertheiner P et al. Ultrasound, computed tomography and endoscopic retrograde cholangiopancreatography in the morphologic diagnosis of pancreatic disease. *Klin Wochenschr*. 1983; 61: 291–296
- [25] Pistolesi GF, Procacci C, Fugazzola C, Marzoli GP, Pederzoli P, Quarta Colosso P. Place of computed tomography in pancreatic disease. Comparison with other radiological methods. *Comput Tomogr* 1981; 5: 115–137
- [26] Almeida RR, Lo GC, Patino M, Bizzo B, Canellas R, Sahani DV. Advances in Pancreatic CT Imaging. *AJR Am J Roentgenol*. 2018; 211: 52–66
- [27] Engjom T, Pham KD, Erchinger F, Haldorsen IS, Gilja OH, Dimcevski G et al. Good Agreement Between Transabdominal and Endoscopic Ultrasound of the Pancreas in Chronic Pancreatitis. *Ultraschall in Med*. 2018
- [28] Pagliari D, Ainora ME, Brizi MG, Cintoni M, Rinninella E, Attili F et al. A new ultrasound score for the assessment and follow-up of chronic pancreatitis: The 'Gemelli USCP score'. *Dig Liver Dis*. 2020; 52: 644–650
- [29] Tanaka S, Fukuda J, Nakao M, Ioka T, Ashida R, Takakura R et al. Effectiveness of Contrast-Enhanced Ultrasonography for the Characterization of Small and Early Stage Pancreatic Adenocarcinoma. *Ultrasound Med Biol*. 2020; 46: 2245–2253
- [30] Dietrich CF, Hocke M. Elastography of the Pancreas, *Current View. Clin Endosc*. 2019; 52: 533–540