



# HHS Public Access

Author manuscript

*AJR Am J Roentgenol.* Author manuscript; available in PMC 2022 August 01.

Published in final edited form as:

*AJR Am J Roentgenol.* 2022 August ; 219(2): 303–313. doi:10.2214/AJR.21.27234.

## Interobserver Agreement for CT and MRI Findings of Chronic Pancreatitis in Children: A Multicenter Ancillary Study Under the INSPPIRE Consortium

**Andrew T. Trout, MD,**

Department of Radiology, Cincinnati Children's Hospital Medical Center

Department of Radiology, University of Cincinnati College of Medicine

Department of Pediatrics, Cincinnati Children's Hospital Medical Center

**Maisam Abu-El-Haija, MD,**

Division of Pediatric Gastroenterology, Hepatology and Nutrition, Cincinnati Children's Hospital Medical Center

Department of Pediatrics, University of Cincinnati College of Medicine

**Sudha A. Anupindi, MD,**

Department of Radiology, The Children's Hospital of Philadelphia, University of Pennsylvania Perelman School of Medicine

**Megan B. Marine, MD,**

Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Riley Hospital for Children

**Michael Murati, MD,**

Department of Radiology, M Health Fairview University of Minnesota Masonic Children's Hospital

**Andrew S. Phelps, MD,**

Department of Radiology and Biomedical Imaging, University of California San Francisco Benioff Children's Hospital

**Mitchell A. Rees, MD,**

Department of Radiology, Nationwide Children's Hospital

**Judy H. Squires, MD,**

Department of Radiology, University of Pittsburgh Medical Center

Department of Pediatric Radiology, UPMC Children's Hospital of Pittsburgh

**Kate M. Ellery, DO,**

---

**Corresponding Author:** Andrew T. Trout, MD, Department of Radiology, MLC 5031, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH, 45229-3026, (p) 513-803-3004, (f) 513-636-8145.

Institution of Origin:

This work was conducted under the INSPPIRE consortium with Cincinnati Children's Hospital Medical Center serving as the coordinating center

Disclosures:

No relevant disclosures

Department of Pediatric Gastroenterology, Hepatology, and Nutrition, UPMC Children's Hospital of Pittsburgh

**Cheryl E. Gariepy, MD,**

Division of Gastroenterology, Hepatology, and Nutrition, Nationwide Children's Hospital

Department of Pediatrics, The Ohio State University College of Medicine

**Asim Maqbool, MD,**

Pediatric Gastroenterology, Hepatology and Nutrition, The Children's Hospital of Philadelphia, University of Pennsylvania Perelman School of Medicine

**Brian A. McFerron, MD,**

Division of Pediatric Gastroenterology, Hepatology & Nutrition. Riley Hospital for Children. Indiana University School of Medicine

**Emily R. Perito, MD,**

Department of Pediatrics, Department of Epidemiology and Biostatistics, University of California San Francisco

**Sarah J. Schwarzenberg, MD,**

Department of Pediatrics, University of Minnesota Masonic Children's Hospital

**Bin Zhang, PhD,**

Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center

Department of Pediatrics, University of Cincinnati College of Medicine

**Dana K. Andersen, MD,**

Division of Digestive Diseases and Nutrition, National Institute of Diabetes and Digestive and Kidney Disease, National Institutes of Health

**Mark E. Lowe, MD, PhD,**

Department of Pediatrics, Washington University School of Medicine

**Aliye Uc, MD**

Division of Gastroenterology, Hepatology, Pancreatology and Nutrition, Stead Family Department of Pediatrics, The University of Iowa

**on behalf of the Consortium for the Study of Chronic Pancreatitis, Diabetes and Pancreatic Cancer (CPDPC), a clinical research consortium funded by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and the National Cancer Institute (NCI)**

## Abstract

**Background:** Imaging findings represent key criteria for diagnosing chronic pancreatitis in children. Understanding radiologists' agreement for imaging findings is critical to standardizing and optimizing diagnostic criteria.

**Objective:** To evaluate the interobserver agreement among experienced pediatric radiologists for subjective, quantitative, and semi-quantitative imaging findings of chronic pancreatitis in children.

**Methods:** In this retrospective study, CT or MRI examinations performed in children with chronic pancreatitis were submitted by six sites participating in the INSPPIRE consortium. One pediatric radiologist from each of the six sites reviewed examinations; three of the radiologists independently reviewed all CT examinations, and the other three radiologists independently reviewed all MRI examinations. Reviewers recorded 13 categorical imaging findings of chronic pancreatitis and measured pancreas thickness and duct diameter. Agreement was assessed using kappa coefficients for the categorical variables and intraclass correlation coefficients (ICC) for the continuous measures.

**Results:** A total of 76 CT and 80 MRI examinations performed in 110 children (mean age, 11.3±4.6 years; 65 girls, 45 boys) were reviewed. For CT, kappa coefficients for categorical findings ranged from -0.01 to 0.81, with relatively high kappa coefficients for parenchymal calcification ( $\kappa=0.81$ ), main pancreatic duct dilation ( $\kappa=0.63$ ), and atrophy ( $\kappa=0.52$ ). ICCs for parenchymal thickness measurements ranged from 0.57 in the pancreas head to 0.80 in the body and tail. ICC for duct diameter was 0.85. For MRI, kappa coefficients for categorical findings ranged from -0.01 to 0.74, with relatively high kappa coefficients for main duct irregularity ( $\kappa=0.74$ ), side branch dilation ( $\kappa=0.70$ ), number of dilated side branches ( $\kappa=0.65$ ), and main duct dilation ( $\kappa=0.64$ ); kappa coefficient for atrophy was 0.52. ICCs for parenchymal thickness measurements ranged from 0.59 in the pancreas head to 0.68 in the tail. ICC for duct diameter was 0.77.

**Conclusion:** Interobserver agreement was fair to moderate for most CT and MRI findings of chronic pancreatitis in children.

**Clinical Impact:** This study highlights challenges for the imaging diagnosis of pediatric chronic pancreatitis. Standardized and/or objective criteria are needed given the importance of imaging in diagnosis.

---

## Introduction

Chronic pancreatitis is a progressive inflammatory disorder characterized by changes in the pancreas that lead to loss of function [1, 2]. Imaging findings are essential diagnostic criteria for pediatric chronic pancreatitis [3]. According to the International Study Group of Pediatric Pancreatitis: In search for a cuRE (INSPPIRE), diagnosis of chronic pancreatitis requires either histologic proof of chronic pancreatitis or imaging findings of chronic pancreatitis plus abdominal pain or exocrine or endocrine pancreatic insufficiency [3]. Imaging, particularly CT and MRI, also plays an important role in non-invasive monitoring and staging of pediatric pancreatitis [4, 5]. Further, imaging findings are increasingly explored as markers of clinically relevant outcomes, including exocrine and endocrine function, in both children and adults [6]. Thus, imaging plays a critical role in the clinical care of children with chronic pancreatitis and in research to improve their outcomes.

Reported findings in chronic pancreatitis include pancreas parenchymal volume loss (i.e., atrophy), parenchymal and intraductal calcifications, abnormal enhancement (usually hypoenhancement or delayed enhancement), main pancreatic duct dilation, pancreatic duct side branch dilation, and main pancreatic duct irregularity or stricture on CT and MRI, as well as parenchymal loss of T1-weighted signal on MRI [5, 7]. Currently, most of

these findings, which have largely been extrapolated from adult studies and criteria, are subjectively assessed.

Given the importance of imaging in clinical care and research related to chronic pancreatitis, it is essential to define interobserver agreement for imaging findings. This knowledge would guide selection of criteria for chronic pancreatitis diagnosis and staging, likely favoring criteria with the highest levels of agreement. It should also inform efforts to refine imaging criteria for future clinical and research efforts. Currently, a paucity of published data are available regarding interobserver agreement on imaging findings in children with chronic pancreatitis.

The purpose of this study was to evaluate the interobserver agreement among experienced pediatric radiologists for subjective, quantitative, and semi-quantitative imaging findings of chronic pancreatitis in children.

## Methods

### Study Design

This was a retrospective HIPAA-compliant multicenter study, conducted as an ancillary study under INSPPIRE with approval from the Study of Chronic Pancreatitis, Diabetes and Pancreatic Cancer (CPDPC) Steering Committee. Study activities were conducted under single institutional review board approval at Cincinnati Children's Hospital Medical Center (hereafter, the central site). The six other sites participating in the study are listed in Table S1 and are all members of the INSPPIRE consortium. Investigators at each site included a pediatric gastroenterologist with an interest and/or expertise in pancreatic disease and a pediatric radiologist with an interest and/or expertise in pancreatic imaging. The pediatric gastroenterologists contributed to patient selection and provided clinical information as correlation for the imaging data. The pediatric radiologists participated in study image reviews. The experience of the participating radiologists is provided in Table S1.

### Sample Size

Sample size was determined based on *a priori* power analysis. We assumed a prevalence of any single finding of chronic pancreatitis of 0.4 and that three radiologists would review each examination with a null hypothesis of  $\kappa=0.6$  and an alternative hypothesis of  $\kappa=0.8$ . Based on these assumptions, and without adjustment for multiple comparisons, we calculated 77 examinations would be needed to achieve 80% power to detect a statistically significant difference using a two-tailed test with  $\alpha=0.05$ . Based on this calculation, we aimed to include a total of approximately 80 CT examinations and 80 MRI examinations in this study.

### Image Submission and Central Review

To achieve the desired study sample size, participating sites (other than the central site) were each requested to submit 14 deidentified 1.5-T or 3-T abdominal MRI examinations and 14 deidentified IV contrast material enhanced abdominal CT examinations performed in patients with known clinical diagnoses of chronic pancreatitis per INSPPIRE criteria [3].

Minimum criteria for submitted MRI examinations were inclusion of axial T1-weighted fat-saturated images and heavily T2-weighted MRCP images. Sites were permitted to submit only one CT examination per patient and only one MRI examination per patient. While sites were permitted to submit both one CT examination and one MRI examination in the same patient, this was not required; when submitting both in the same patient, no criteria were provided regarding a minimum or maximum interval between the two modalities. Sites were instructed not to submit imaging examinations from patients with prior pancreatic surgery. No other specific criteria were provided regarding selection of imaging examinations for submission. The imaging examinations had been performed locally at participating sites according to standard institutional protocols and thus varied in technique beyond the previously noted criteria for submission. While all sites submitting images were participants in INSPPIRE, the submitted examinations were not required to have been performed in patients who were enrolled in INSPPIRE. Participating sites also submitted demographic and clinical data extracted from the patient's medical record for each submitted imaging examination.

All submitted examinations were centrally reviewed by the study principal investigator affiliated with the central site (A.T.T., a pediatric radiologist with 8 years of postfellowship experience) to confirm the presence of one or more findings of chronic pancreatitis (e.g., parenchymal atrophy, parenchymal T1-weighted signal loss [MRI], parenchymal or duct calcifications, main duct dilation, main duct irregularity, side branch dilation). During this central review, the principal investigator also evaluated the examinations for all of the findings that were subsequently reviewed by the participating radiologists. Examinations demonstrating one or more findings of chronic pancreatitis and thus meeting inclusion criteria were uploaded to an online image portal (Ambra; Ambra Health, New York, NY) for subsequent review by the radiologists. Examinations without findings of chronic pancreatitis based on central review, as well as examinations in which the pancreas could not be assessed due to the presence of large fluid collection(s), were replaced with additional examinations from participating sites or from the central site to achieve the target sample size for each modality. The study principal investigator did not participate in the subsequent image review to quantify interobserver agreement.

## Image Review

One board-certified pediatric radiologist from each of the six participating sites (other than the central site) participated in independent image review. Three radiologists reviewed only CT examinations (A.S.P., M.A.R., J.H.S.), and three radiologists reviewed only MRI examinations (S.A.A., M.B.M., M.M.). Each examination was reviewed by the three radiologists performing the image review for the given modality, and each radiologist reviewed every examination of the modality to which they were assigned. Reviewers were blinded to each other's findings, to the findings of the initial review by the principal investigator, and to examination details (e.g., performing site and clinical information). Reviewers evaluated 13 categorical subjective features (12 on both CT and MRI, 1 on CT only, 1 on MRI only) (Table 1). Reviewers also performed quantitative measurements of pancreas thickness (CT and MRI), pancreatic duct diameter (CT and MRI), and signal intensity on unenhanced T1-weighted images of pancreas, spleen, and paraspinous muscle

(MRI). Prior to image review, the reviewers received guidance regarding location and orientation of pancreas thickness measurements. According to the guidance, pancreas thickness measurements were performed perpendicular to the surfaces of the pancreas in the head/uncinate, neck, body and tail, as previously described by Trout et al. (Fig. 1) [8]; the sequence to use for measuring pancreas thickness and pancreatic duct diameter on MRI was not specified. The signal intensity measurements were performed using ROIs placed in the three measured structures on a single axial slice on an unenhanced T1-weighted sequence. The ROI size was at the discretion of the individual radiologist, though the three ROIs were requested to be of a similar size for each patient. Mean values for each ROI were submitted, and signal intensity ratios between the pancreas and two reference tissues (i.e., pancreas-to-spleen and pancreas-to-muscle ratios) were calculated centrally.

### Statistical Analysis

Categorical variables were expressed using frequencies and percentages. Means and SD were used to summarize continuous variables. Frequencies of findings as recorded by the central reviewer were summarized to serve as an estimate of the relative frequency of each finding in the study sample. Fleiss' kappa coefficient for categorical variables and intra-class correlation coefficients (ICC) for continuous variables were used to assess inter-observer agreement. The percentage of cases for which all three readers agreed was also calculated for categorical variables. Agreement was quantified among the three independent site radiologists as a primary outcome. Agreement between individual radiologists and the central reviewer was quantified as a secondary outcome (supplemental material). To explore the effect of the presence of acute pancreatitis on interobserver agreement, interobserver agreement analysis was also performed for subgroups of patients with acute pancreatitis and patients without acute pancreatitis (supplemental material). Similarly, to explore the effect of IV contrast material on interobserver agreement for MRI, interobserver agreement analyses were performed for subgroups of patients imaged without IV contrast material and patients imaged without and with IV contrast material (supplemental material).

Kappa coefficients were interpreted as [9]: 0.20, slight; 0.21 to 0.40, fair; 0.41 to 0.60, moderate; 0.61 to 0.80, substantial, and 0.81 to 1, almost perfect. Intraclass correlation coefficients were interpreted as [10]: 0 to 0.49, poor; 0.50 to 0.74, moderate; 0.75 to 0.90, good; and >0.90, excellent. To quantify mean variability among observers in continuous measurements, the SD of measurements from all observers for each patient was calculated, and the mean of the SD was determined. A p value less than .05 was considered statistically significant for all inference testing. All analyses were performed using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA).

### Results

Submitted imaging examinations were acquired between July 2009 and November 2019 for CT and between December 2019 and August 2020 for MRI. Based on central review of submitted images by the principal investigator, two submitted examinations were excluded due to absence of imaging findings of chronic pancreatitis. An additional 10 examinations were excluded due to the presence of fluid collections replacing a large

portion of the pancreas. Replacement examinations for excluded examinations included 11 examinations from the central site for CT. After the exclusions and replacements, a total of 76 CT examinations and 80 MRI examinations were reviewed by the participating site radiologists. The target sample size of 80 examinations was not reached for CT due to insufficient availability of replacement examinations. The 156 reviewed examinations had been performed in 110 unique patients; 46 patients had both CT and MRI examinations included in the analysis. All 80 CT examinations were performed with IV contrast material. Thirty-nine (49%) MRI examinations were performed with IV contrast material. A total of 26 MRI examinations were performed at 1.5 T, and 54 were performed at 3 T. The study sample is summarized in Table 2. The 110 unique patients had a mean age at the time of the first included imaging examination of  $11.3 \pm 4.6$  years; 65 were female, and 45 were male.

### CT Findings

Frequencies of CT findings as recorded by the central reviewer are reported in Table 3. According to central review, 51% of examinations had findings of acute pancreatitis at the time of imaging. Agreement among the independent observers for the presence of acute pancreatitis was good ( $\kappa=0.65$ ).

At central review, the most common subjective features of chronic pancreatitis were pancreas atrophy (61%) and duct dilation (68%) (Fig. 2). Interobserver agreement among the three reviewers for the subjective and quantitative findings is reported in Table 3. Interobserver agreement for each individual reviewer with the central reviewer is reported in Table S2. Kappa coefficients for the categorical features ranged from  $-0.01$  to  $0.81$ . Among the features, agreement among reviewers was relatively high for the presence of parenchymal calcifications ( $\kappa=0.81$ , almost perfect), pancreatic duct dilation ( $\kappa=0.63$ , substantial), and atrophy ( $\kappa=0.52$ , moderate) (Fig. 3, Fig. 4).

Agreement among reviewers for measurements of pancreas parenchymal thickness ranged from moderate in the head (ICC=0.57) to good in the body and tail (both ICC=0.80). The mean SD among reviewers, as a measure of variability, was  $4.2 \pm 2.9$  mm in the head,  $1.9 \pm 1.3$  mm in the neck,  $2 \pm 1.4$  mm in the body, and  $1.4 \pm 1.1$  mm in the tail. Agreement for measurement of pancreatic duct diameter was good (ICC=0.85) with a mean SD among reviewers of  $0.9 \pm 0.8$  mm.

For the subgroup analyses assessing agreement separately for patients without versus with acute pancreatitis as determined by central review, the 95% CIs for the measures of interobserver agreement overlapped for all findings except for measurements of pancreas parenchymal thickness in the head (ICC=0.76 for no acute pancreatitis; 0.45 for acute pancreatitis) and of main pancreatic duct diameter (ICC=0.74 for no acute pancreatitis; 0.89 for acute pancreatitis), and for presence of anomalous pancreaticobiliary junction ( $\kappa=1$  for no acute pancreatitis;  $\kappa=-0.02$  for acute pancreatitis) (Table S3).

### MRI Findings

Frequencies of MRI findings as recorded by the central reviewer are reported in Table 4. According to central review, 51% of examinations had findings of acute pancreatitis at the



time of imaging. Agreement among the independent observers for the presence of acute pancreatitis was fair ( $\kappa=0.42$ ).

At central review, the most common subjective features of chronic pancreatitis were any severity of loss of T1-weighted signal (63%), pancreas atrophy (65%), main pancreatic duct irregularity (69%), and main pancreatic duct dilation (81%) (Fig. 5). Interobserver agreement among the three independent reviewers for the subjective and quantitative findings, as well as T1 signal intensity ratio, is reported in Table 4. Interobserver agreement for each individual reviewer with the central reviewer is reported in Table S4. Kappa coefficients for the categorical features ranged from  $-0.01$  to  $0.74$ . Among the features, agreement among the reviewers was relatively high for main pancreatic duct irregularity ( $\kappa=0.74$ , substantial), the presence of dilated pancreatic duct side branches ( $\kappa=0.70$ , substantial), number of dilated pancreatic duct side branches ( $\kappa=0.65$ ), and main pancreatic duct dilation ( $\kappa=0.64$ , substantial) (Fig. 6). Agreement among reviewers for the presence of atrophy was moderate ( $\kappa=0.52$ ) (Fig. 7).

Agreement among reviewers for measurements of pancreas parenchymal thickness was moderate at all locations (ICC= $0.53$ – $0.68$ ). The mean SD among reviewers was  $3.9 \pm 3.6$  mm in the head,  $2.0 \pm 1.5$  mm in the neck,  $2.5 \pm 1.5$  mm in the body, and  $2.0 \pm 1.6$  mm in the tail. Agreement for measurement of pancreatic duct diameter was good (ICC= $0.77$ ) with a mean SD among reviewers of  $0.9 \pm 0.9$  mm.

For the subgroup analyses assessing agreement separately for patients with versus without acute pancreatitis as determined by central review, the 95% CIs for the measures of interobserver agreement overlapped for all findings except for measurements of pancreas-to-spleen T1 signal intensity ratio (ICC= $0.44$  for no acute pancreatitis;  $0.83$  for acute pancreatitis) (Table S5).

For the subgroup analyses assessing agreement separately for patients imaged with MRI without versus with and without IV contrast material, the 95% CIs for the measures of interobserver agreement overlapped for all findings except for presence of anomalous pancreaticobiliary junction ( $\kappa=0.37$  for without IV contrast material;  $\kappa=-0.01$  for without and with IV contrast material) (Table S6).

## Discussion

In this study of cross-sectional imaging in children with chronic pancreatitis from centers in the multi-center INSPPIRE study, we have shown that most subjective imaging features of chronic pancreatitis have only fair to moderate agreement among experienced pediatric radiologists. For CT, the highest level of agreement was for pancreas calcifications, for which agreement was almost perfect. For MRI, the highest levels of agreement were for duct changes of chronic pancreatitis (including main duct irregularity) and side branch dilation, for which agreement was substantial.

A limited number of prior studies have explored agreement for imaging features of chronic pancreatitis, and all prior studies to our knowledge were performed in adult populations. In a study based on the PROCEED cohort of adult patients, Tirkes et al. reported agreement



between local and central reviewing radiologists to be only moderate for a composite CT score ( $\kappa=0.56$ ) and substantial for a composite MRI score ( $\kappa=0.68$ ) [11]. Individual imaging findings were not evaluated. Razek et al. evaluated agreement for individual CT findings defined according to reporting standards of the CPDPC [12, 13]. Their study of imaging from 47 adults reviewed by two radiologists showed substantial to excellent agreement for all assessed CT findings ( $\kappa=0.71-0.87$ ). Finally, Lisitskaya et al. performed a study with a design similar to our present study in which 80 CT and 80 MRI examinations performed in adult patients with chronic pancreatitis were reviewed by two radiologists [14]. Their study had similar findings as in our study, with almost perfect agreement for detection of calcifications by CT ( $\kappa=0.87$ ) and fair to moderate agreement for other CT findings. Regarding MRI findings, with the exception of main pancreatic duct dilation, Lisitskaya et al. reported weaker agreement for duct findings of chronic pancreatitis ( $\kappa=0.32-0.66$ ) than in our study. Factors that may contribute to the observed differences in agreement between our study and prior studies include differences in disease severity between children and adults, differences in common causes of chronic pancreatitis between children and adults, and methodological differences (e.g., two vs three reviewers). Adult patients would be expected to have more severe or advanced disease, potentially leading to more conspicuous imaging findings.

Agreement was somewhat better for quantitative measures of pancreas parenchymal thickness and duct diameter than for subjective imaging findings. Agreement was good for duct diameter measurements on both CT and MRI, moderate to good for measurements of parenchymal thickness on CT, and moderate for measurements of parenchymal thickness on MRI. For CT, agreement for parenchymal thickness measurements was better in the body and tail than in the head, an observation that may have been driven by poor agreement in the head for patients with acute pancreatitis. In patients without acute pancreatitis, agreement was similar across all segments of the pancreas. Lisitskaya et al. also evaluated agreement for pancreatic duct and parenchyma thickness measurements in adults [14]. Compared with our study, agreement for all such measurements was in general better in their study, with ICCs of approximately 0.75 for duct diameter and 0.80–0.94 for pancreas parenchymal thickness. This discrepancy could relate to the presence of only two reviewers in their study. Additionally, examinations performed for chronic pancreatitis may be of better image quality in adults, given greater ability by adults to comply with breath hold instructions during imaging. Further, the relatively greater volume of intraabdominal fat in adults may facilitate identification of the margins of the pancreas.

While interobserver agreement for pancreas thickness measurements have not been reported in children with pancreatitis to our knowledge, prior studies have reported such measurements in healthy children. For example, Aydin et al. reported ICC values of 0.65 to 0.87 for MRI-based measurements of healthy pancreas parenchymal thickness by two reviewers [15]. In addition, Trout et al. reported ICC values of 0.52 to 0.70 for CT-based measurements of healthy pancreas parenchymal thickness by three reviewers [8].

In addition to the subjective findings and quantitative measures, we also explored agreement for pancreas signal intensity ratios on MRI as a semiquantitative measure of the extent of pancreas parenchyma loss of T1-weighted signal. T1-weighted signal loss is recognized as

a finding of chronic pancreatitis in both children and adults and has been linked to exocrine dysfunction [5, 12, 16]. Agreement for pancreas signal intensity ratio was better when the pancreas was compared to muscle (ICC=0.87) than when compared to spleen (ICC=0.68). The reason for this difference is uncertain. Nonetheless, this finding may be relevant for the optimal application of signal intensity ratios in clinical care and research. To our knowledge, interobserver agreement for pancreas signal intensity ratios on T1-weighted MRI has not been previously assessed in children or adults.

While CT and MRI examinations were reviewed by different sets of independent reviewers, agreement on duct findings of chronic pancreatitis was in general better for MRI than for CT (with the exception of subjective assessment of main pancreatic duct diameter and measurement of main duct diameter). This observation is expected as the higher soft tissue contrast resolution of MRI, augmented by T2-weighted sequences and heavily T2-weighted MRCP sequences, increases the conspicuity of the ducts. For this reason, MRI is the imaging modality of choice when non-invasively assessing the pancreatic duct [4, 5]. Agreement for pancreas parenchymal thickness was somewhat better for CT than for MRI for all pancreas segments except for the head. This finding may relate to the lack of specification given to the reviewers of which MRI sequence to use for measuring pancreas thickness, introducing a potential source of variability to the measurements.

The findings of our study and prior studies provide guidance for advancement and use of CT and MRI features of chronic pancreatitis in pediatric clinical practice and research. First, in general, subjective assessments of disease may be suboptimal for clinical and research use based on limited interobserver agreement. Second, if subjective criteria are used, specific definitions are needed [14], and those definitions need to be developed specifically for children. Third, quantitative and semiquantitative measures of disease may provide better agreement in comparison with subjective criteria, but methods of measurement need to be standardized.

Our study has limitations. First, we had one fewer CT examination than necessary to meet our target sample size defined by *a priori* power calculation given the inability to identify a replacement examination. Additionally, included CT examinations were performed over a much longer time interval than the included MRI examinations. Both of these limitations reflect the much less frequent use of CT than MRI in children with chronic pancreatitis. Second, findings of acute pancreatitis were identified in approximately 50% of imaging examinations. While recurrent acute pancreatitis is a common indication for imaging in children with known chronic pancreatitis, the presence of parenchymal edema and swelling is known to impact the imaging conspicuity of findings of chronic pancreatitis [5]. Subgroup analyses suggest that the presence of acute pancreatitis had small impact on interobserver agreement. We observed differences in agreement, based on lack of overlap between 95% CIs, only for measurement of head thickness and measurement of duct diameter on CT and measurement of pancreas-to-spleen T1 signal intensity ratio on MRI. In the presence of acute pancreatitis, agreement for head thickness measurements by CT was weaker, possibly due to the presence of edema, and agreement for duct diameter measurements was better, possibly due to larger duct diameter in patients with acute pancreatitis; however, the association of duct diameter and acute pancreatitis was not evaluated. For MRI, agreement

for pancreas-to-spleen signal intensity ratio was better in the presence of acute pancreatitis. The cause of this finding is uncertain. Imaging assessment of chronic pancreatitis would optimally be performed during a quiescent interval to prevent acute inflammation from obscuring relevant findings. Third, submitted imaging examinations had been performed according to local protocols and were thus not standardized. Finally, the reviewers did not receive formal training prior to reviewing examinations for purposes of this investigation, and specific definitions were not provided for some subjective findings (e.g. atrophy, duct dilation, severity of T1-weighted signal loss). Such steps for achieving standardization, if implemented, may have improved interobserver agreement [14]. Defining criteria, however, would optimally be based on knowledge of normal, which changes for some features during childhood, as well as on knowledge of thresholds relevant to diagnosing disease. Such data are becoming available but remain limited in children. On the other hand, the pediatric radiologists participating in the independent review were experienced in imaging of chronic pancreatitis, which may have favorably impacted interobserver agreement.

In conclusion, interobserver agreement for imaging findings of chronic pancreatitis in children was fair to moderate for most findings on CT and MRI. For both modalities, duct findings showed the highest levels of agreement. Agreement for subjective findings of atrophy was moderate, and agreement for parenchymal thickness (a quantitative measure of atrophy) depended on the location of measurement, with the best agreement observed in the pancreatic body. Our results highlight challenges in the interpretation of imaging of pediatric chronic pancreatitis and suggest the need for standardized and/or objective criteria. Additional studies of imaging findings of chronic pancreatitis in large samples of children are needed and ideally will include serial assessment of imaging and detailed correlation with clinical data.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Funding Sources:

Research reported in this publication was supported by Quad Cities Pediatrics Fund for Pediatric Gastroenterology and National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under award numbers U01 DK108334. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## REFERENCES:

1. Schwarzenberg SJ, Bellin M, Husain SZ, et al. Pediatric chronic pancreatitis is associated with genetic risk factors and substantial disease burden. *J Pediatr* 2015; 166:890–896 e891 [PubMed: 25556020]
2. Schwarzenberg SJ, Uc A, Zimmerman B, et al. Chronic Pancreatitis: Pediatric and Adult Cohorts Show Similarities in Disease Progress Despite Different Risk Factors. *J Pediatr Gastroenterol Nutr* 2019; 68:566–573 [PubMed: 30897605]
3. Morinville VD, Husain SZ, Bai H, et al. Definitions of pediatric pancreatitis and survey of present clinical practices. *J Pediatr Gastroenterol Nutr* 2012; 55:261–265 [PubMed: 22357117]
4. Trout AT, Anupindi SA, Freeman AJ, et al. Naspghan and the Society for Pediatric Radiology Joint Position Paper on Non-Invasive Imaging of Pediatric Pancreatitis: Literature Summary and Recommendations. *J Pediatr Gastroenterol Nutr* 2020;

5. Trout AT, Ayyala RS, Murati MA, et al. Current State of Imaging of Pediatric Pancreatitis: AJR Expert Panel Narrative Review. *AJR Am J Roentgenol* 2021;1–13
6. Tirkes T, Yadav D, Conwell DL, et al. Magnetic resonance imaging as a non-invasive method for the assessment of pancreatic fibrosis (MINIMAP): a comprehensive study design from the consortium for the study of chronic pancreatitis, diabetes, and pancreatic cancer. *Abdom Radiol (NY)* 2019; 44:2809–2821 [PubMed: 31089778]
7. Chavhan GB, Babyn PS, Manson D, Vidarsson L. Pediatric MR cholangiopancreatography: principles, technique, and clinical applications. *Radiographics* 2008; 28:1951–1962 [PubMed: 19001651]
8. Trout AT, Preet-Singh K, Anton CG, et al. Normal pancreatic parenchymal thickness by CT in healthy children. *Pediatr Radiol* 2018; 48:1600–1605 [PubMed: 29951835]
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159–174 [PubMed: 843571]
10. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016; 15:155–163 [PubMed: 27330520]
11. Tirkes T, Shah ZK, Takahashi N, et al. Inter-observer variability of radiologists for Cambridge classification of chronic pancreatitis using CT and MRCP: results from a large multi-center study. *Abdom Radiol (NY)* 2020; 45:1481–1487 [PubMed: 32285180]
12. Tirkes T, Shah ZK, Takahashi N, et al. Reporting Standards for Chronic Pancreatitis by Using CT, MRI, and MR Cholangiopancreatography: The Consortium for the Study of Chronic Pancreatitis, Diabetes, and Pancreatic Cancer. *Radiology* 2019; 290:207–215 [PubMed: 30325281]
13. Razek A, Elfar E, Abubacker S. Interobserver agreement of computed tomography reporting standards for chronic pancreatitis. *Abdom Radiol (NY)* 2019; 44:2459–2465 [PubMed: 30955070]
14. Lisitskaya MV, Olesen SS, Svarc P, et al. Systematic approach for assessment of imaging features in chronic pancreatitis: a feasibility and validation study from the Scandinavian Baltic Pancreatic Club (SBPC) database. *Abdom Radiol (NY)* 2020; 45:1468–1480 [PubMed: 32170335]
15. Aydin S, Fatihoglu E, Karavas E, Kantarci M. Normal pancreatic thickness values in healthy children: an MRI study. *Pancreatology* 2021;
16. Tirkes T, Fogel EL, Sherman S, et al. Detection of exocrine dysfunction by MRI in patients with early chronic pancreatitis. *Abdom Radiol (NY)* 2017; 42:544–551 [PubMed: 27660281]

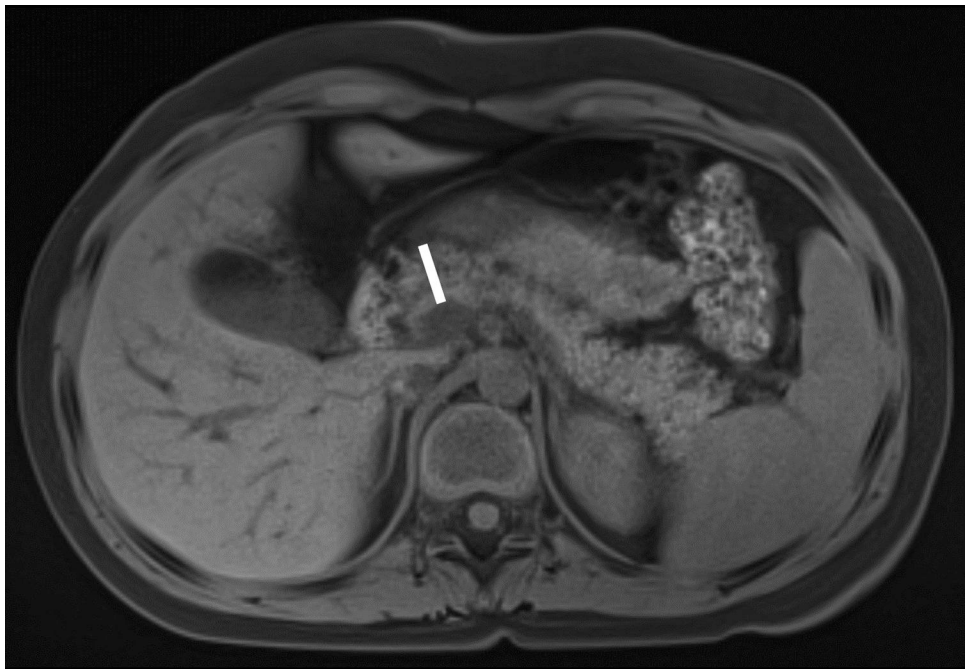
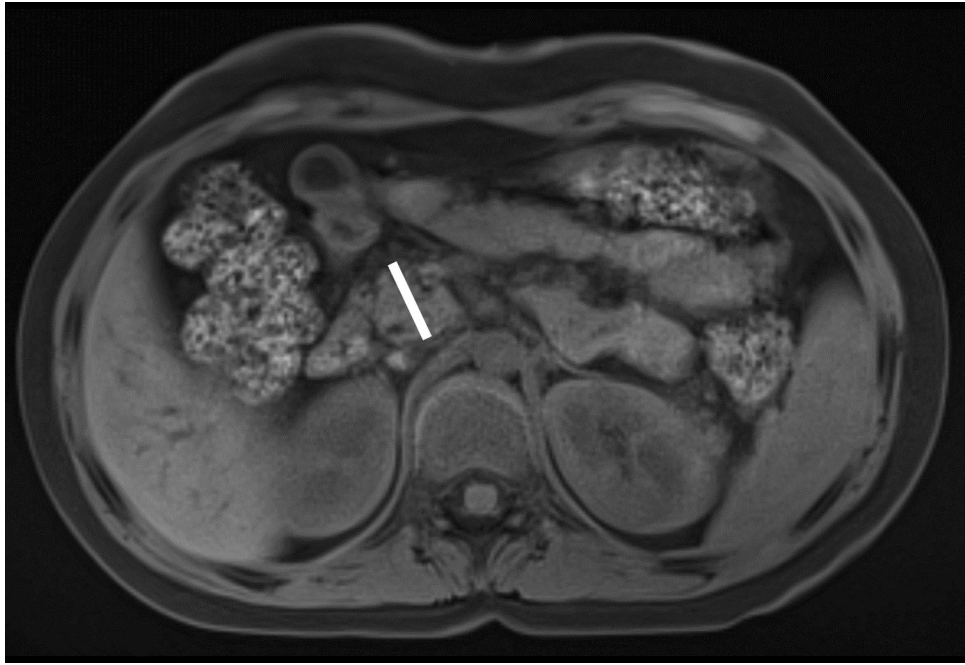
### HIGHLIGHTS

**Key Finding:**

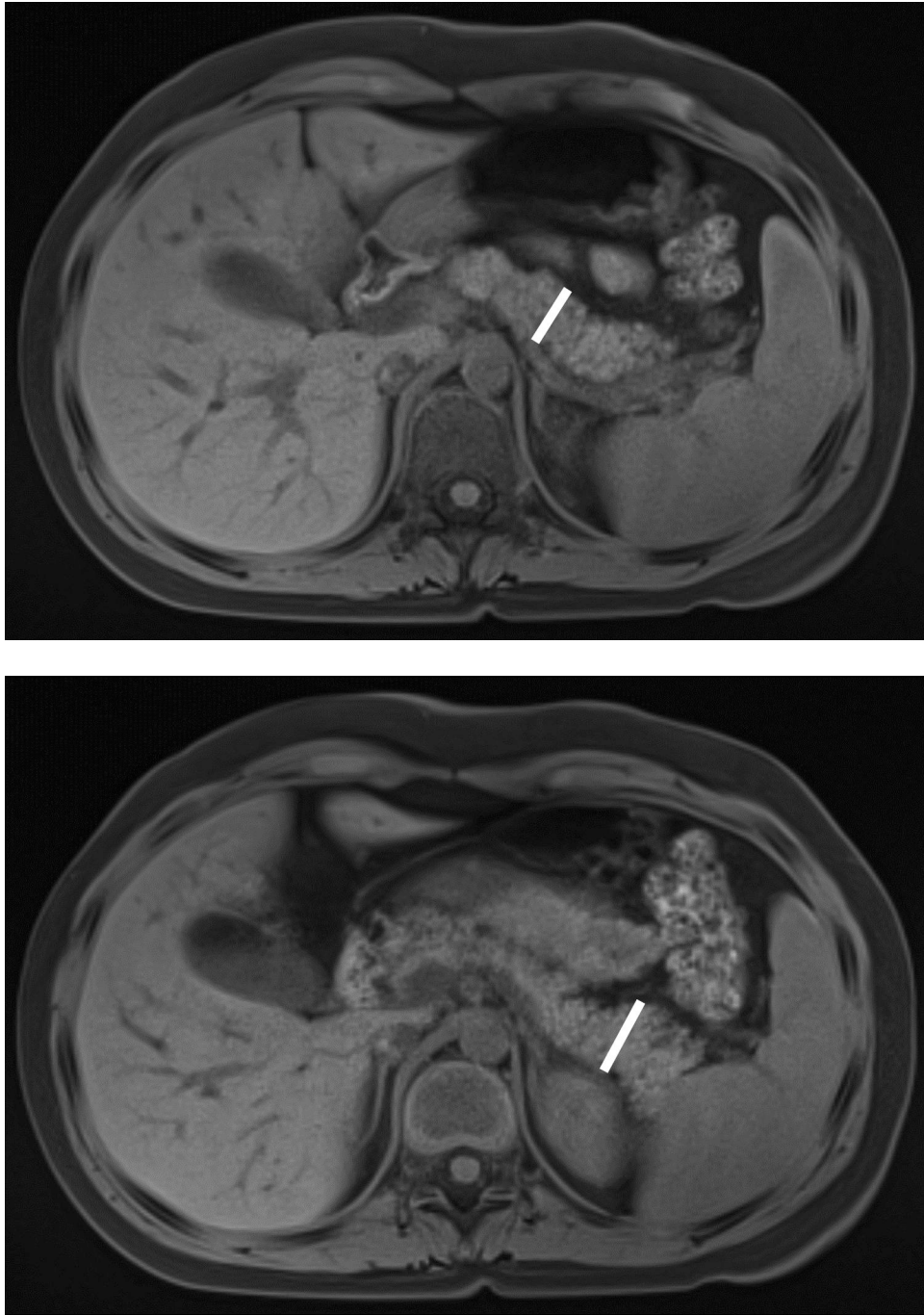
In this study of six sites in the INSPPIRE consortium, interobserver agreement for findings of chronic pancreatitis in children was relatively high for CT for the presence of parenchymal calcifications ( $\kappa=0.81$ ), pancreatic duct dilation ( $\kappa=0.63$ ), and atrophy ( $\kappa=0.52$ ); and for MRI for main and side branch pancreatic duct findings ( $\kappa=0.64-0.74$ ).

**Importance:**

Interobserver agreement for findings of chronic pancreatitis in children was generally fair to moderate, highlighting challenges in interpretation and need for standardized and/or objective criteria.

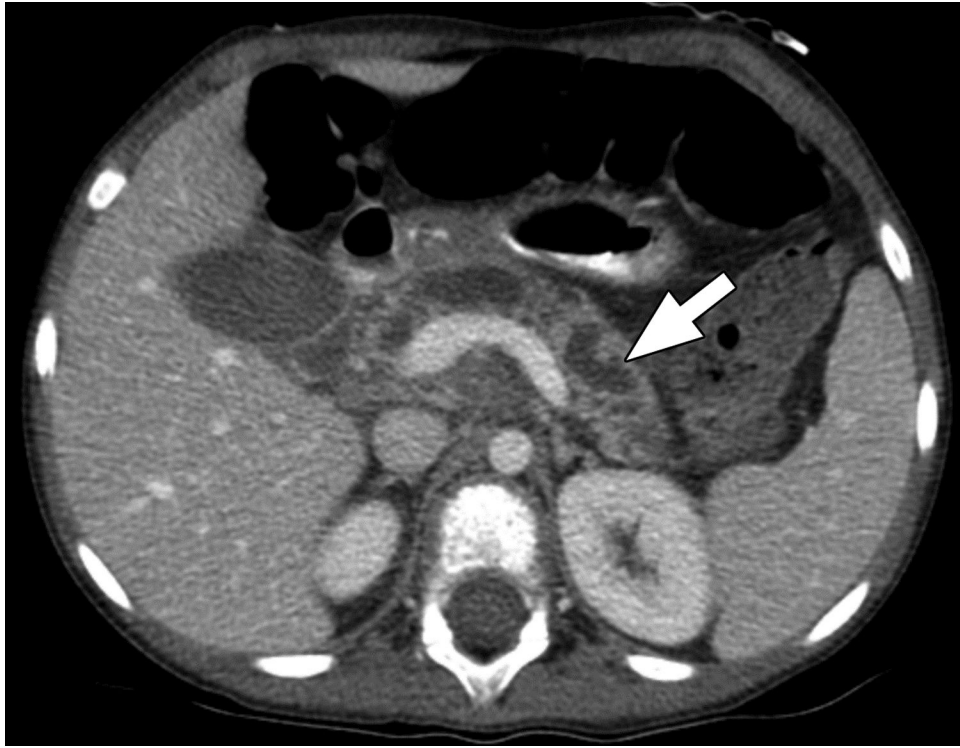






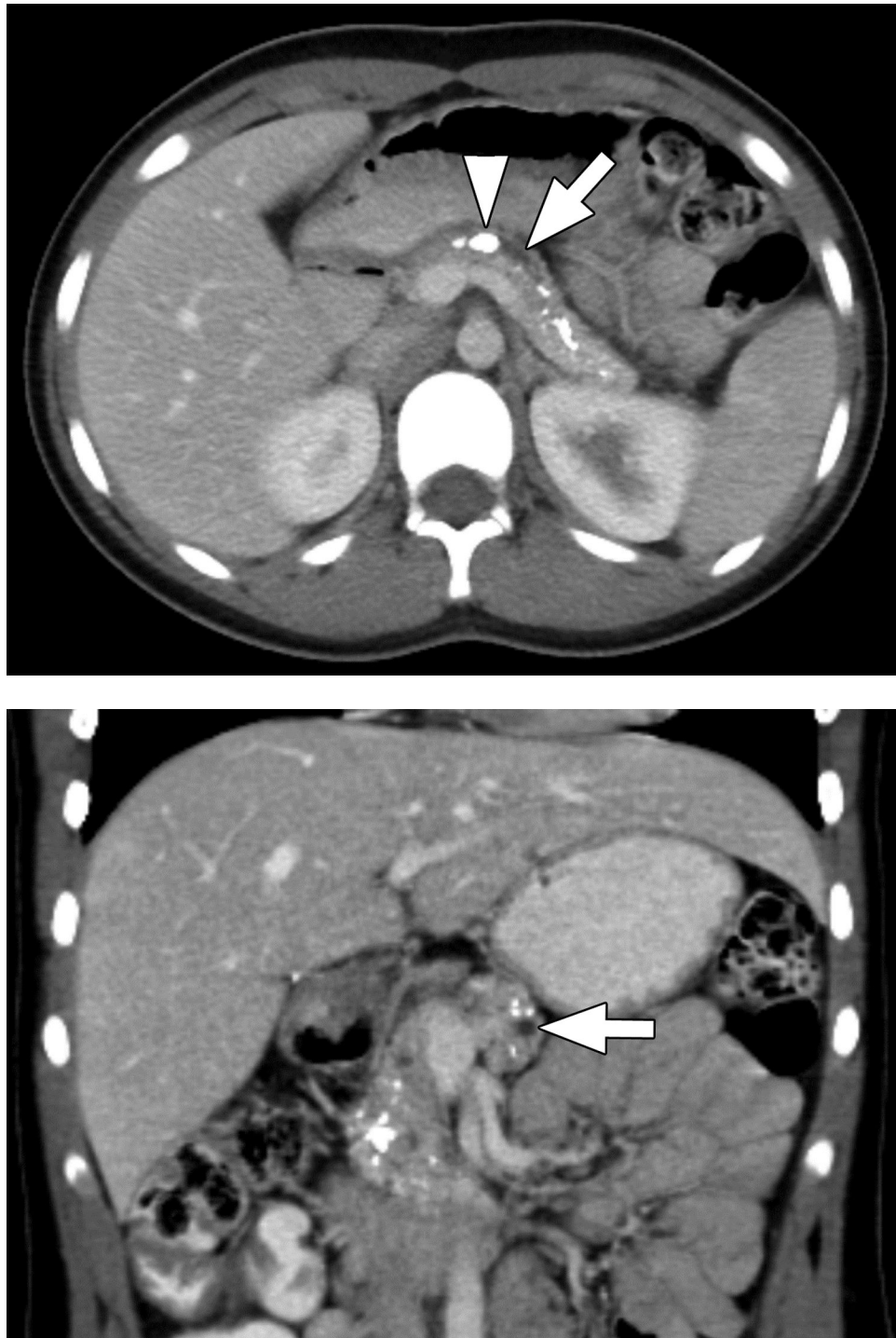
**Figure 1 –.** Axial T1-weighted fat-saturated gradient recalled echo MR images in 10-year-old girl with chronic pancreatitis, showing measurement of pancreas parenchymal thickness. As previously described [8], measurements were made perpendicular to the surfaces of the pancreas to the right of the superior mesenteric vein and anterior to the inferior vena cava for the head (A), anterior to the superior mesenteric vein for the neck (B), anterior or to the left of the abdominal aorta for the body (C), and 1.5 cm to the right of the tip for the pancreatic tail (D).





**Figure 2 –.**

Axial contrast-enhanced CT images in 7-year-old girl with chronic pancreatitis. The main pancreatic duct is severely dilated (arrow), and the surrounding pancreatic parenchyma is diffusely thinned (atrophy). The central reviewer and all three site radiologists agreed on the presence of main pancreatic duct dilation and pancreas atrophy.



**Figure 3 –.** Parenchymal atrophy and calcifications on CT in 13-year-old girl with chronic pancreatitis. (A) axial and (B) coronal reformatted images from CT performed with IV and oral contrast material show the pancreas to be small with extensive parenchymal calcifications. Duct calcifications are also present (arrowhead in A). Two of three reviewers agreed on the presence of atrophy, a finding for which agreement among reviewers on CT was moderate

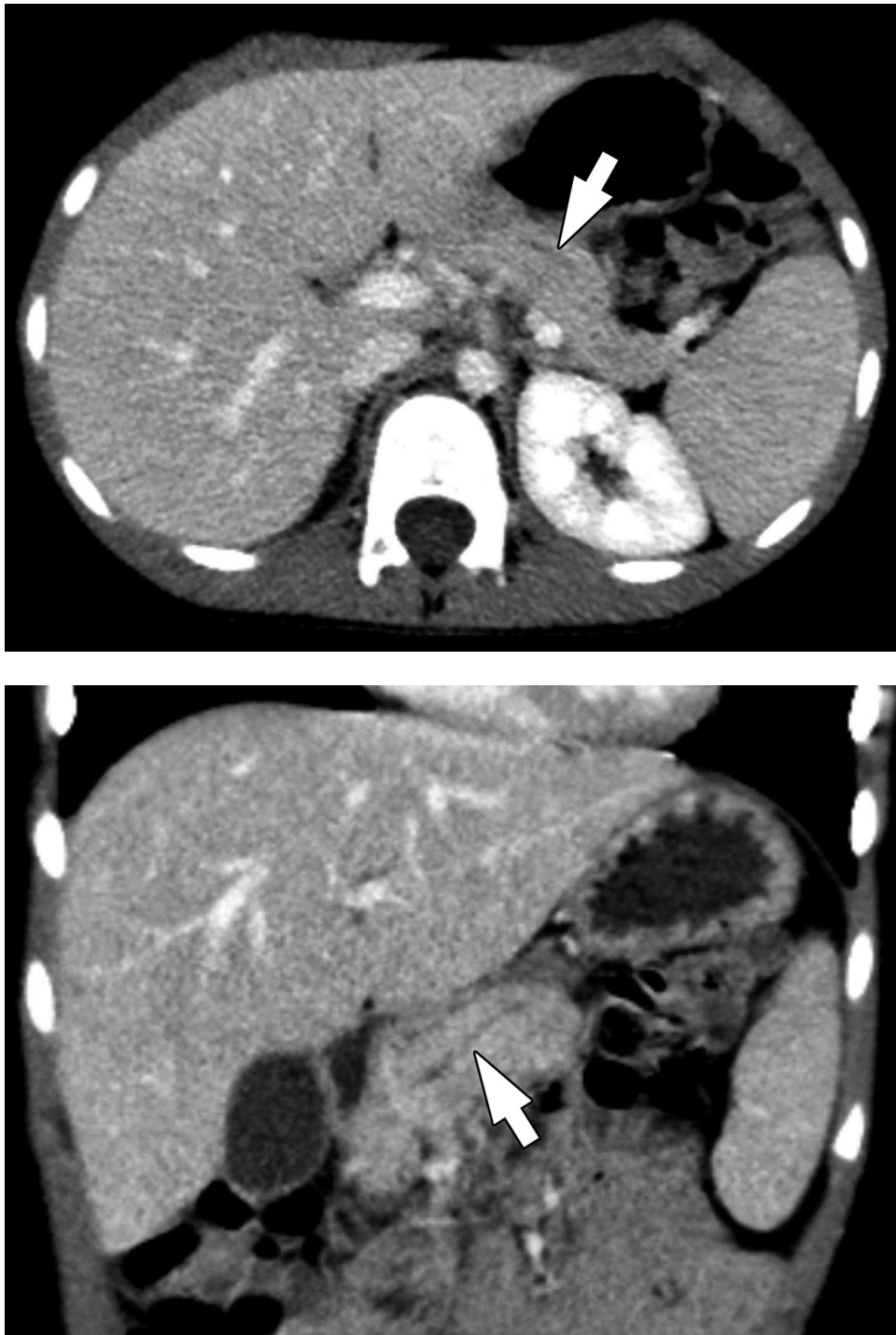
( $\kappa=0.52$ ). All three reviewers agreed on the presence of parenchymal calcifications and main duct dilation (arrows), findings for which agreement among reviewers on CT was almost perfect ( $\kappa=0.81$ ) and substantial ( $\kappa=0.63$ ), respectively.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 4 –.** Duct dilation on CT in 6-year-old boy with chronic pancreatitis. (A) Axial and (B) coronal reformatted images from CT performed with IV contrast material show segmental visualization of the main pancreatic duct (arrow). One of three reviewers described main duct dilation to be present. Agreement among reviewers on CT for main duct dilatation was substantial ( $\kappa=0.63$ ).





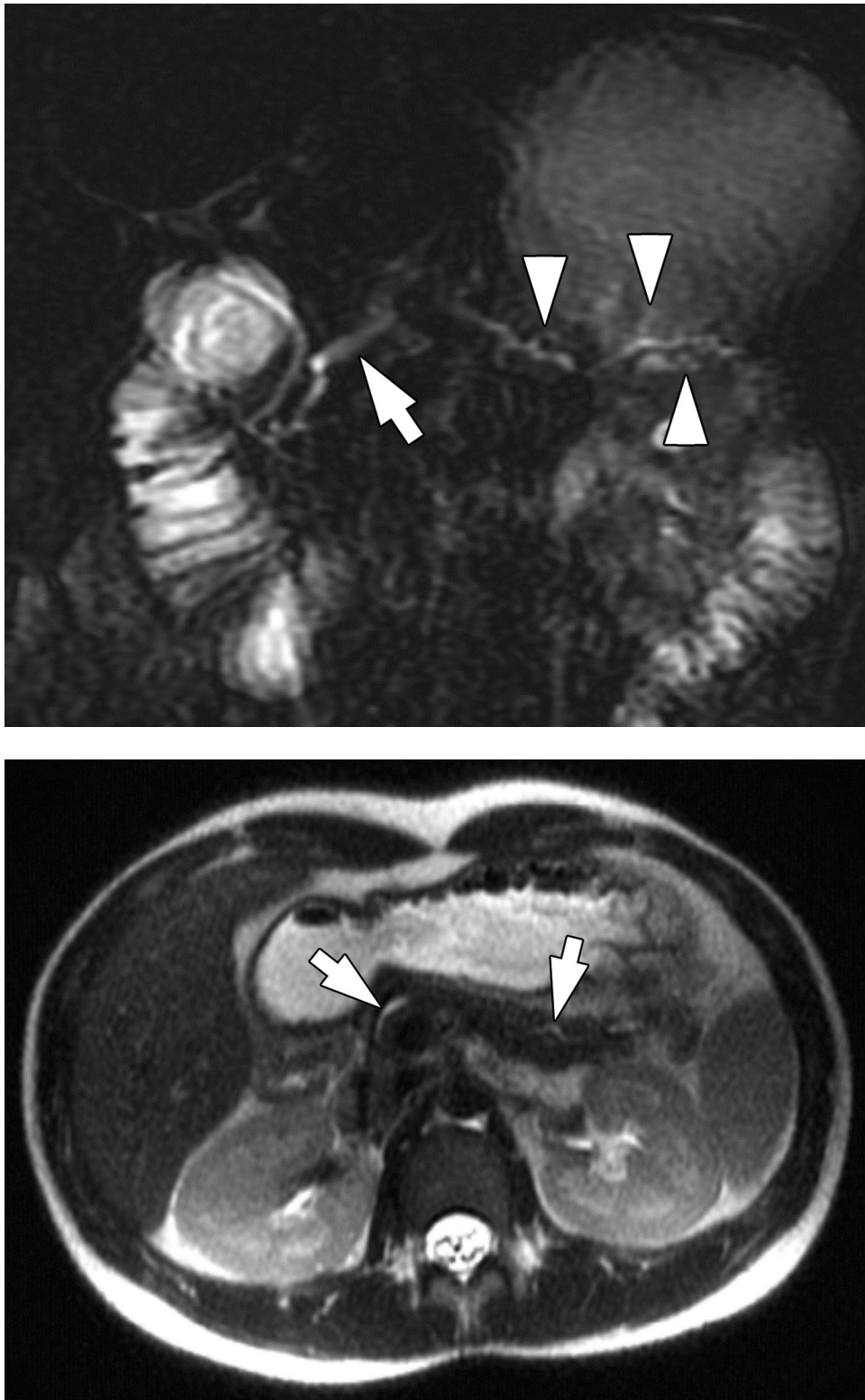
**Figure 5 –.**  
MR images from 14-year-old girl with chronic pancreatitis. (A) Axial T1-weighted 3D fast spoiled gradient-echo (FSPGR) water image shows pancreas atrophy (arrow) and heterogeneous loss of T1-weighted signal. (B) Axial T2-weighted SSFSE image shows pancreas atrophy (arrow), main pancreatic duct dilation, and irregularity (arrowhead). The central reviewer and all three site radiologists agreed on the presence of main pancreatic duct dilation, main pancreatic duct irregularity, pancreas atrophy, and loss of T1-weighted signal.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 6 –.** Side branch dilation on MRI in 16-year-old boy with chronic pancreatitis. (A) Coronal 3D T2-weighted MRCP maximum intensity projection image and (B) axial T2-weighted image



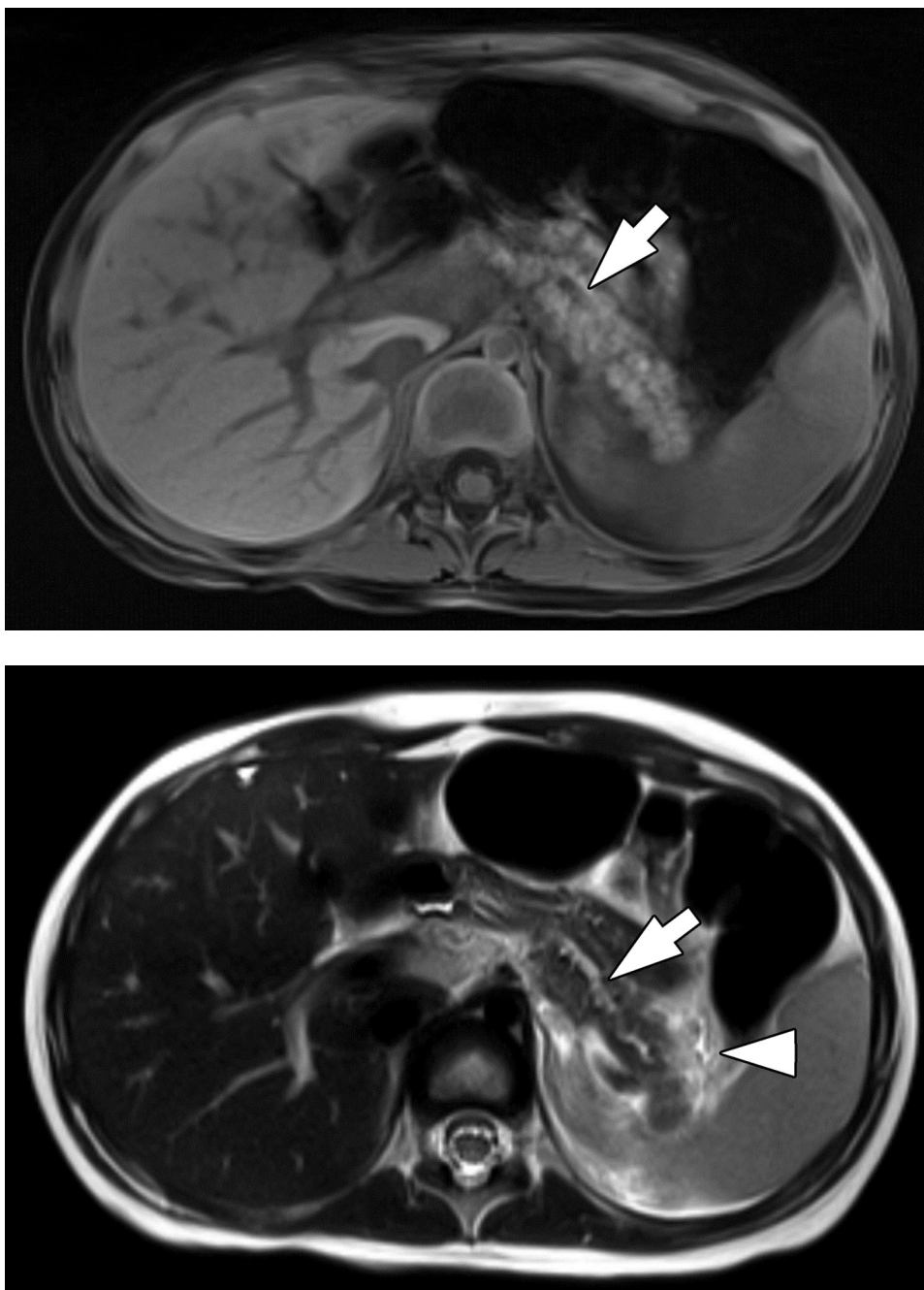
show dilated main pancreatic duct (arrows) with dilated side branches in the pancreas body and tail (arrowheads). Two of three reviewers described side branch dilation to be present, with one describing 1–2 side branches to be dilated and one describing 3 side branches to be dilated. All three reviewers agreed on the presence of main pancreatic duct dilation. Agreement among three reviewers on MRI for main pancreatic duct dilation was substantial ( $\kappa=0.64$ ) and for dilated pancreatic duct side branches was substantial ( $\kappa=0.70$ ). One of three reviewers also described atrophy to be present. Agreement among three reviewers on MRI for pancreas atrophy was moderate ( $\kappa=0.52$ ).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 7 –.** Parenchymal atrophy on MRI in 4-year-old girl with chronic pancreatitis. (A) Axial radial T1-weighted gradient recalled echo fat-saturated image and (B) axial T2-weighted image show diffuse thinning and increased lobulation of the pancreas. The main pancreatic duct is also dilated (arrows). T1-weighted signal is preserved and is higher than liver. Two of three reviewers reported the presence of atrophy, and all reviewers reported the presence of main duct dilation and absence of loss of T1-weighted signal. Agreement among three reviewers on MRI for atrophy was moderate ( $\kappa=0.52$ ), for main pancreatic duct dilation was substantial ( $\kappa=0.64$ ), and for loss of T1-weighted signal was fair ( $\kappa=0.37$ ). On T2-weighted

imaging, edema is present adjacent to the pancreatic tail (arrowhead). Two of three reviewers reported the presence of acute pancreatitis, a finding for which agreement among three reviewers on MRI was moderate ( $\kappa=0.42$ ).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1 –**

Categorical subjective imaging findings evaluated by radiologist reviewers. The modality column indicates for which modality the assessment was performed. The scoring column indicates how each finding was scored.

<b>Finding</b>	<b>Modality</b>	<b>Scoring</b>
Pancreas atrophy	MRI, CT	Present, Absent
Parenchymal calcifications	CT	Present, Absent
Pancreas divisum	MRI, CT	Present, Absent
Anomalous pancreaticobiliary junction	MRI, CT	Present, Absent
Pancreatic duct dilation	MRI, CT	Present, Absent
Pancreatic duct irregularity	MRI, CT	Present, Absent
Pancreatic duct stricture (focal narrowing with upstream dilation)	MRI, CT	Present, Absent
Dilation of pancreatic duct side branches	MRI, CT	Present, Absent
Number of dilated pancreatic duct side branches	MRI, CT	0, 1–2, 3
Intraductal (pancreatic duct) filling defects	MRI, CT	Present, Absent
Loss of T1-weighted signal	MRI	Present, Absent
Parenchymal enhancement	MRI, CT	Normal / Hypoenhancing / Hyperenhancing / Heterogeneous
Acute pancreatitis	MRI, CT	Present, Absent
Gallstones	MRI, CT	Present, Absent

**Table 2 –**

## Characteristics of study sample

Characteristic	Value
Unique patients	110 (100)
Patients with CT only	30 (27)
Patients with MRI only	34 (31)
Patients with both CT and MRI	46 (42)
Age <sup>a</sup> (y), mean $\pm$ SD	11.3 $\pm$ 4.6
Interval between imaging examinations <sup>b</sup> (y), mean $\pm$ SD	0.75 $\pm$ 1.05
Sex	
Male	45 (41)
Female	65 (59)

Unless otherwise indicated, values reported as number of patients, with percentage in parentheses

<sup>a</sup>In patients who had both CT and MRI examinations included in analysis, age was calculated at the time of the earlier of the two examinations.

<sup>b</sup>Calculated among the 46 patients for whom both a CT examination and an MRI examination were included in the independent multireader review

Table 3 –

Interobserver agreement of CT findings.

Finding	Central Reviewer Findings	$\kappa$ (95% CI)	ICC (95% CI)	Three-Rater Agreement
Pancreas atrophy	46/76 (61)	0.52 (0.39, 0.65)	NA	49/76 (64)
Pancreas thickness (mm), mean $\pm$ SD				
Head	25.6 $\pm$ 6.4		0.57 (0.47, 0.67)	
Neck	11.5 $\pm$ 3.4	NA	0.73 (0.65, 0.80)	NA
Body	17.1 $\pm$ 5.1		0.80 (0.74, 0.85)	
Tail	16.7 $\pm$ 5.3		0.80 (0.74, 0.85)	
Parenchymal calcifications	16/76 (21)	0.81 (0.68, 0.94)	NA	69/76 (91)
Pancreas divisum	11/76 (14)	0.33 (0.20, 0.46)	NA	66/76 (87)
Anomalous pancreatocobiliary junction	3/76 (4)	-0.01 (-0.14, 0.12)	NA	74/76 (97)
Pancreatic duct dilation	52/76 (68)	0.63 (0.50, 0.76)	NA	57/76 (75)
Pancreatic duct diameter (mm), mean $\pm$ SD	3.8 $\pm$ 2.9	NA	0.85 (0.79, 0.89)	NA
Pancreatic duct irregularity	30/76 (39)	0.40 (0.27, 0.53)	NA	42/76 (55)
Pancreatic duct stricture	22/76 (29)	0.21 (0.08, 0.34)	NA	42/76 (55)
Dilation of pancreatic duct side branches	18/76 (24)	0.30 (0.17, 0.43)	NA	43/76 (57)
Number of dilated pancreatic duct side branches	0: 58/76 (76) 1-2: 17/6 (1) 3: 17/76 (22)	0.23 (0.06, 0.41)	NA	37/76 (49)
Intraductal (pancreas) filling defects	10/76 (13)	0.32 (0.19, 0.45)	NA	72/76 (95)
Parenchymal enhancement	Normal: 61/76 (80) Hypoenhancing: 11/76 (14) Hyperenhancing: 0/76 (0) Heterogeneous: 4/76 (5)	0.41 (0.24, 0.57)	NA	41/76 (54)
Acute pancreatitis	39/76 (51)	0.65 (0.52, 0.78)	NA	56/76 (74)

Finding	Central Reviewer Findings	$\kappa$ (95% CI)	ICC (95% CI)	Three-Rater Agreement
Gallstones	3/76 (4)	0.07 (-0.06, 0.20)	NA	68/76 (89)

Central reviewer findings reported as numerator and denominator with percentage in parentheses, unless otherwise reported. Kappa coefficients and ICC presented for the three reviewers. Three-rater agreement reports the fraction of examinations for which all three reviewers agreed on the presence or absence of a finding for binary features or on the specific category for findings with three or four categories, expressed as numerator and denominator with percentage in parentheses.

ICC = intraclass correlation coefficient; NA = not applicable



**Table 4 –**

Interobserver agreement for MRI findings<sup>a</sup>.

Finding	Central Reviewer Findings	κ (95% CI)	ICC (95% CI)	Three-Rater Agreement
Pancreas atrophy	52/80 (65)	0.52 (0.40, 0.65)	NA	52/80 (65)
Pancreas thickness (mm), mean ± SD				
Head	22.7±5.7		0.59 (0.49, 0.68)	
Neck	11.2±3.2	NA	0.53 (0.43, 0.63)	NA
Body	16.3±5.0		0.62 (0.52, 0.70)	
Tail	16.3±4.9		0.68 (0.59, 0.76)	
Pancreas divisum	15/80 (19)	0.49 (0.36, 0.62)	NA	67/80 (84)
Anomalous pancreaticobiliary junction	1/80 (1)	0.32 (0.19, 0.44)	NA	76/80 (95)
Pancreatic duct dilation	65/80 (81)	0.64 (0.51, 0.77)	NA	60/80 (75)
Pancreatic duct diameter (mm), mean ± SD	4.2±2.4	NA	0.77 (0.71, 0.83)	NA
Pancreatic duct irregularity	55/80 (69)	0.74 (0.62, 0.87)	NA	66/80 (83)
Pancreatic duct stricture	33/80 (41)	0.47 (0.34, 0.59)	NA	52/80 (65)
Dilation of pancreatic duct side branches	48/80 (60)	0.70 (0.57, 0.82)	NA	62/80 (78)
Number of dilated pancreatic duct side branches	0: 32/80 (40) 1-2: 6/80 (8) 3: 42/80 (53)	0.65 (0.52, 0.78)	NA	51/80 (64)
Intraductal (pancreas) filling defects	19/80 (24)	0.49 (0.36, 0.61)	NA	60/80 (75)
Loss of T1-weighted signal	Any: 50/80 (63) Mild: 15/80 (19) Moderate: 27/80 (34) Severe: 8/80 (10)	0.37 (0.26, 0.47)	NA	27/80 (34)
T1 signal intensity ratio				
Pancreas-to-spleen	1.1±0.3	NA	0.68 (0.59, 0.75)	NA
Pancreas-to-muscle	1.1±0.4		0.87 (0.83, 0.90)	

Finding	Central Reviewer Findings	$\kappa$ (95% CI)	ICC (95% CI)	Three-Rater Agreement
Parenchymal enhancement	Normal: 19/38 (50) Hypoenhancing: 9/38 (24) Hyperenhancing: 1/38 (0) Heterogeneous: 10/38 (26)	0.43 (0.22, 0.65)	NA	50/80 (63)
Acute pancreatitis	41/80 (51)	0.42 (0.29, 0.55)	NA	56/80 (70)
Gallstones	4/80 (5)	-0.01 (-0.13, 0.12)	NA	78/80 (98)

Central reviewer findings reported as numerator and denominator with percentage in parentheses, unless otherwise reported. Kappa coefficients and ICC presented for the three reviewers. Three-rater agreement reports the fraction of examinations for which all three reviewers agreed on the presence or absence of a finding for binary features or on the specific category for findings with three or four categories, expressed as numerator and denominator with percentage in parentheses.

<sup>a</sup>38/80 examinations performed without and with IV contrast material

ICC = intraclass correlation coefficient; NA = not applicable