G.16 Liver disease

Review question: What is the effectiveness of ultrasound scanning to detect clinically important cystic fibrosis related liver disease?

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Full citation	Sample size n=70 CF patients	Tests Reference test	Methods	Results	Limitations QUADAS 2 checklist

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Fagundes, E. D. T., Silva, R. A. P., Roquete, M. L. V., Penna, F. J., Reis, F. J. C., Goulart, E. M. A., Duque, C. G., Validation of the Williams ultrasound scoring system for the diagnosis of liver disease in cystic fibrosis, Jornal de Pediatria, 80, 380-386, 2004 Ref Id 354000 Country/ies where the study was carried out Brazil Study type Prospective cohort study Aim of the study To describe the hepatic abnormalities viewed in the ultrasound scans of CF patients at the Cystic Fibrosis Outpatients Clinic at the Hospital das Clínicas of UFMG, to compare these ultrasound findings	Characteristics Mean age, years (SD): 10.9 (6.4) 60% male 14.3% met the clinical and/or biochemical criteria for liver disease Inclusion Criteria -Confirmed CF diagnoses Exclusion Criteria Other causes of liver disease, such as Wilson's disease, hepatitis B and C, deficiency of alpha-1-antitrypsin and auto-immune hepatitis	Clinical and biochemical criteria. The clinical examination was considered abnormal when the presence of a palpable spleen and/or hepatomegaly, defined as the presence of a palpable liver more than 2.5 cm below the right costal margin (RCM), of firm consistency. Abnormal biochemistry was defined as a significant and persistent increase, of at least 1.5 times the upper limit of the reference range, of at least two of the enzymes aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (AP) or gamma-glutamyltranspeptidase (GGT), for a period of more than 6 months. Index tests Williams ultrasound score: normal	Setting: CF outpatient clinic at a Brazilian university Seventy cystic fibrosis patients were followed prospectively and underwent clinical, biochemical and ultrasound examinations. The ultrasound findings were compared to the results of the clinical and biochemical examinations. Clinical and biochemical criteria were used as the gold standard for the validation of the Williams ultrasound score. We calculated the sensitivity, specificity, and positive and negative predictive values for the Williams score. The patients were divided into two groups: normal (score = 3) or abnormal (score > 3) ultrasound examination.	Williams US score versus clinical and/or biochemical criteria for detection of CFLD n=70 True positive=5* False positive=5* False negative=5* True negative=55* Sensitivity=50 (95% CI: 22.0-75.1)* Specificity=91.7 (95% CI: 87.0-95.8)* Positive LR= 6.0 (95% CI: 1.70-18.07)* Negative LR= 0.55 (95% CI: 0.26-0.90)* AUROC=NR *Calculated by the NGA technical team from data reported in the article NR=not reported LR = likelihood ratio	Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? N/A 2.A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN Reference Standard

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
with biochemical and clinical criteria and validate the Williams score for the diagnosis of CF-associated liver disease. Study dates 1999-2000 Source of funding Not reported		ultrasound results (score = 3) or abnormal (score > 3). Patients underwent the hepatobiliary ultrasound examination at the Radiology Service of the Hospital das Clínicas at UFMG. All examinations were performed by the same ultrasound operator with no regard to the clinical and biochemical situation of the patients. The apparatus employed was from the Siemens Prima line, a multi-frequency (2.6 to 5.0 MHz) Sonoline Prima, with convex probe.Abnormalities in the echogenicity of the hepatic parenchyma and edge were noted as was periportal fibrosis, in accordance with the scoring devised by Williams et al. Signs suggestive of steatosis, the presence of ascites and collateral portal system damage were noted in addition to			Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3.A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3.B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW RISK Flow and Timing Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4.A Could the patient flow have introduced bias? LOW RISK Other information None.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		measurements for the liver, spleen and gallbladder taken with the electronic pachymeter. The right lobe of the liver was measured from the phrenic cupola to its lower edge, at the level of the right hemiclavicular line, to the right of the gallbladder bed and the left lobe, in turn, from the phrenic cupola to the lower edge, at the level of the sagittal line. The longitudinal axis of the spleen was measure at the level of the medial axillary line and the anterior-posterior along the left flank. Reference values for liver and spleen measurements for the different age groups were taken from a study by Konus et al.			
Full citation Karlas, T., Neuschulz, M., Oltmanns, A., Guttler, A., Petroff, D., Wirtz, H., Mainz,	Sample size 55 adults with CF 14 with CFLD Characteristics Total study cohort/without	Tests Reference test Cystic fibrosis-related liver disease was defined if at least 2 of the following	Methods Adult CF patients were prospectively investigated at presentation to the pulmonary outpatient	Results TE versus published criteria for detection of CFLD n=49 True positive=6*	Limitations QUADAS 2 checklist Patient selection

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
J. G., Mossner, J., Berg, T., Troltzsch, M., Keim, V., Wiegand, J., Non-invasive evaluation of cystic fibrosis related liver disease in adults with ARFI, transient elastography and different fibrosis scores, PLoS ONE [Electronic Resource], 7, e42139, 2012 Ref Id 354030 Country/ies where the study was carried out Germany Study type Case-control study Aim of the study Evaluate transient elastography (TE), acoustic radiation force impulse imaging (ARFI), and fibrosis indices for CFLD detection. Study dates April-Dec 2010 Source of funding None	CFLD/with CFLD/CFLD without cirrhosis/CFLD with cirrhosis Male, n: 31/24/7/4/3 Age, year, mean (SD): 31.9(8.8)/32.9(9.0)/29.0(8.0)/29.6(7.8)/28.3(8.9) Inclusion Criteria Adult CF patients Exclusion Criteria Patients with pregnancy, age < 18 years, and liver transplantation	conditions were present on at least 2 consecutive examinations spanning a 1-year period [6,7]: (1) Ultrasound confirmed hepatomegaly; (2) elevated serum liver enzyme levels of ALT, AST, AP, or GGT; (3) ultrasound abnormalities other than hepatomegaly (i.e., increased, heterogeneous echogenicity, nodularity, irregular margins, splenomegaly). An ultrasonographic pattern of simple liver steatosis did not represent a diagnostic criterion. In case of distinct ultrasonographic signs of liver cirrhosis (i.e. coarse nodularity, presence of portal hypertension and rarefication of peripheral portal veins) and clinical signs (e.g. esophageal varices, splenomegaly) of liver cirrhosis CFLD	clinic for clinical routine examinations. Patients with pregnancy, age 18 years, and liver transplantation were not included. Patients underwent conventional upper abdomen ultrasound evaluation, elastography and blood tests (alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (AP), bilirubin, gammaglutamyltransferase (GGT), blood count, INR, albumin, creatinine, and cholesterol) at the same day. Fasting for at least three hours was required prior to examination, however exceptions were permitted when clinically required. Previous ultrasound reports, recent pulmonary function tests (time span < 6 months), and results of previous routine blood tests were collected from clinical records.	False positive=1* False negative=8* True negative=34* Sensitivity=42.9 (95% CI: 22.6-49.6)* Specificity=97.1 (95% CI: 89.0-99.8)* Positive LR= 15.0 (95% CI: 2.06-328.3)* Negative LR= 0.59 (95% CI: 0.51-0.87)* AUROC=0.68 (95% CI: 0.53-0.80) APRI versus published criteria for detection of CFLD n=55 True positive=12* False positive=12* False negative=2* True negative=29* Sensitivity=85.7 (95% CI: 60-97.4)* Specificity=70.7 (95% CI: 62.0-74.7)* Positive LR= 2.93 (95% CI: 1.58-3.86)* Negative LR= 0.20 (95% CI: 0.04-0.65)* AUROC=0.82 (95% CI: 0.69-0.91)	Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? No Did the study avoid inappropriate exclusions? Yes 1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? N/A 2.A Could the conduct or interpretation of the index test have introduced bias? UNCLEAR 2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN Reference Standard

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		patients were classified as cirrhotics. Index tests -Transient elastography (TE): all subjects were examined in a supine position immediately after ARFI measurement. TE was performed in a right intercostal space in resting respiratory position. 10 valid measurements were taken according to the manufacturer's recommendation (M probe). Measurements were performed by experienced operators (TK, VK, MN, MT). Patients with an interquartile range (IQR).median value/3 or a success rate below 60% were considered as invalid and excluded from further analysis. -AST/Platelets-Ratio-Index (APRI) -Forns' score was calculated according to the formula: score=7.811-3:131 x	TE and ARFI were performed in 55 adult CF patients. In addition, AST/Platelets-Ratio-Index (APRI), and Forns' score were calculated. Healthy probands and patients with alcoholic liver cirrhosis served as controls	FORNS versus published criteria for detection of CFLD n=55 True positive=13* False positive=16* False negative=1* True negative=25* Sensitivity=92.9 (95% CI: 67.8-99.6)* Specificity=61.0 (95% CI: 52.4-63.3)* Positive LR= 2.38 (95% CI: 1.43-2.71)* Negative LR= 0.12 (95% CI: 0.006-0.61)* AUROC=0.79 (95% CI: 0.65-0.89) TE versus published criteria for detection of CFLD cirrhosis n=14 True positive=6* False positive=2* False negative=0.5** True negative=6* Sensitivity=92.3 (95% CI: 56.2-100)* Specificity=75 (95% CI: 45.7-81.2)* Positive LR= 3.69 (95% CI: 1.04-5.33)*	Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Unclear 3.A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3.B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW RISK Flow and Timing Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? No 4.A Could the patient flow have introduced bias? LOW RISK Other information None.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		platelet count (109 /l)+ 0.781 x ln GGT (UI/l) + 3.467 x ln age (years)- 0.014xcholesterol (mg/dl)		Negative LR= 0.10 (95% CI: 0-0.96)* AUROC=0.88 (95% CI: 0.59-0.99) APRI versus published criteria for detection of CFLD cirrhosis n=14 True positive=5* False positive=1* False negative=1* True negative=7* Sensitivity=83.3 (95% CI: 45.0-98.5)* Specificity=87.5 (95% CI: 58.8-98.9)* Positive LR= 6.67 (95% CI: 1.09-88.5)* Negative LR= 0.19 (95% CI: 0.02-0.94)* AUROC=0.88 (95% CI: 0.59-0.99) FORNS versus published criteria for detection of CFLD cirrhosis n=14 True positive=4* False positive=0.5** False negative=2* True negative=8* Sensitivity=66.7 (95% CI: 30.1-75.0)*	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Specificity=94.1 (95% CI: 68.3-100)* Positive LR= 11.3 (95% CI: 0.95-6684670)* Negative LR= 0.35 (95% CI: 0.25-1.02)* AUROC=0.85 (95% CI: 0.57-0.98) *Calculated by the NGA technical team from data reported in the article **0.5 person was added by the NGA technical team to calculate likelihood ratios and 95% confidence intervals. LR = likelihood ratio	
Full citation Kitson, M. T., Kemp, W. W., Iser, D. M., Paul, E., Wilson, J. W., Roberts, S. K., Utility of transient elastography in the non-invasive evaluation of cystic fibrosis liver disease, Liver International, 33, 698-705, 2013 Ref Id 354034	Sample size n=50 adults Characteristics All (n=50)/CFLD (n=25)/No CFLD (n=25) Age, years, mean (SD): 23.3 (9.6)/30.5 (9.3)/34.1 (9.8) Male, %: 46/44/48 Diabetes, %: 40/52/28 UDCA, %: 58/88/28	Tests Reference test Diagnosis of CFLD was established according to established criteria if least two of the following conditions on consecutive examinations spanning a one-year period were present: (i) Hepatomegaly and/or splenomegaly confirm ed by ultrasound, (ii) abnormal serum liver	Methods Setting: large CF referral centre in Australia Fifty adult patients with CF were prospectively studied: 25 with CFLD and 25 without CFLD. The presence of CFLD and portal hypertension (PHT) was assessed according to strict established criteria based on serial biochemistry and	Results LSM ≥6.8kPa using TE versus recent guidelines for detection of CFLD n=50 True positive=19* False positive=2* False negative=6* True negative=23* Sensitivity=76 (95% CI: 61.6-82.5)* Specificity=92 (95% CI: 77.6-98.5)* Positive LR= 9.5 (95% CI: 2.75-55.6)*	Limitations QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? No Did the study avoid inappropriate exclusions? Yes 1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Country/ies where the study was carried out Australia Study type Case-control study Aim of the study To evaluate LSM as a diagnostic tool in adults with CFLD. Study dates 2009-2010 Source of funding Not reported	Inclusion Criteria Adult patients with CF and CFLD. Exclusion Criteria Other causes of abnormal liver enzyme levels	enzyme levels, consisting of elevation above the upper limit of normal of 2 of the following: ALT, AST, GGT, (iii) ultrasound abnormalities other than hepatomegaly (increased, heterogeneous echogenicity, nodularity, irregular margins; splenomegaly; presence of porto-systemic collatoeral veins; ascites). Index tests -Transient elastography (TE): Liver stiffness by TE was evaluated usingFibroScan® apparatus and mediam (M) probe by 3 experienced operators. All readings were taken from the right lobe of the liver with an appropriate site for LSM readings identified in the mid-axillary line using conventional US. The median value of 10 successful	imaging. All patients underwent LSM; APRI, Hepascore(®) and Forns score were calculated. This is a prospective case-control study of 50 adults with CF. Control subjects were unmatched patients with CF, but without evidence of liver disease. Cases were patients with CFLD. Optimal LSM values for the prediction of CFLD, PHT and varices were identified by estimating sensitivity and specificity for various cut offs.	Negative LR= 0.26 (95% CI: 0.18-0.50)* AUROC=0.87 (95% CI: 0.77-0.98) LSM ≥ 8.9 kPa using TE versus recent guidelines for detection of portal hypertension for all patients n=50 True positive=7* False positive=4* False negative=1* True negative=38* Sensitivity=87.5 (95% CI: 51.4-99.3)* Specificity=90.5 (95% CI: 83.6-92.7)* Positive LR= 9.19 (95% CI: 3.14-13.66)* Negative LR= 0.14 (95% CI: 0.01-0.58)* AUROC=0.96 (95% CI: 0.92-1.00) LSM ≥ 8.9 kPa using TE versus recent guidelines for detection of portal hypertension for CFLD patients n=25 True positive=7* False positive=4*	the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? N/A 2.A Could the conduct or interpretation of the index test have introduced bias? UNCLEAR 2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN Reference Standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Unclear 3.A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3.B Is there concern that the target condition as defined by the reference standard does

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		acquisitions, expressed in kPA, was taken as representative of LSM. TE was considered valid if 10 successful measurements with a success rate ≥ 60% and an interquartile range (IQR)/median ratio ≤ 30% of the median were obtained. -AST/Platelets-Ratio-Index (APRI) performed at baseline		False negative=1* True negative=13* Sensitivity=87.5 (95% CI: 52.9-99.3)* Specificity=76.5 (95% CI: 60.2-82.0)* Positive LR= 3.7 (95% CI: 1.33-5.53)* Negative LR= 0.16 (95% CI: 0.01-0.78)* AUROC=0.91 (95% CI: 0.79-1.00) APRI ≥ 0.49 versus recent guidelines for detection of portal hypertension for all patients n=50 True positive=7* False positive=3* False negative=1* True negative=39* Sensitivity=87.5 (95% CI: 52.0-99.3)* Specificity=92.9 (95% CI: 86.1-95.1)* Positive LR= 12.3 (95% CI: 3.74-20.3)* Negative LR= 0.14 (95% CI: 0.01-0.56)* AUROC=0.97 (95% CI: 0.93-1.00)	not match the review question? LOW RISK Flow and Timing Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? No 4.A Could the patient flow have introduced bias? LOW RISK Other information

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				APRI ≥ 0.49 versus recent guidelines for detection of portal hypertension for CFLD patients n=25 True positive=7* False positive=1* False negative=1* True negative=16* Sensitivity=87.5 (95% CI: 54.8-98.9)* Specificity=94.1 (95% CI: 78.7-99.5)* Positive LR= 14.9 (95% CI: 2.6-189.4)* Negative LR= 0.13 (95% CI: 0.01-0.58)* AUROC=0.98 (95% CI: 0.93-1.00) Forns ≥ 6.8 versus recent guidelines for detection of portal hypertension for all patients n=50 True positive=7* False positive=6* False negative=1* True negative=36* Sensitivity=87.5 (95% CI: 50.7-99.3)* Specificity=85.7 (95% CI: 78.7-88.0)*	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Positive LR= 6.13 (95% CI: 2.38-8.26)* Negative LR= 0.15 (95% CI: 0.01-0.63)* AUROC=0.93 (95% CI: 0.85-1.00) Forns ≥ 6.8 versus recent guidelines for detection of portal hypertension for CFLD patients n=25 True positive=7* False positive=3* False negative=1* True negative=14* Sensitivity=87.5 (95% CI: 53.2-99.3)* Specificity=82.4 (95% CI: 66.2-87.9)* Positive LR= 5.0 (95% CI: 1.6-8.2)* Negative LR= 0.15 (95% CI: 0.01-0.71)* AUROC=0.93 (95% CI: 0.82-1.00) LSM ≥ 8.9 kPa using TE versus recent guidelines for detection of oesophageal varices for all patients n=23 True positive=6*	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				False positive=4* False negative=0* True negative=13* Sensitivity=100 (95% CI: 57.8-100)* Specificity=76.5 (95% CI: 61.6-76.5)* Positive LR= 4.25 (95% CI: 1.51-4.25)* Negative LR= 0 (95% CI: 0-0.69)* AUROC=0.91 (95% CI: 0.78-1.00) APRI ≥ 0.49 versus recent guidelines for detection of oesophageal varices for all patients n=23 True positive=6* False positive=1* False negative=0* True negative=16* Sensitivity=100 (95% CI: 60.0-100)* Specificity=94.1(95% CI: 80.0-94.1)* Positive LR= 17.0 (95% CI: 3.0-17.0)* Negative LR= 0 (95% CI: 0-0.50)* AUROC=0.99 (95% CI: 0.96-1.00)	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				APRI ≥ 0.49 versus recent guidelines for detection of oesophageal varices for CFLD patients n=13 True positive=6* False positive=0.5** False negative=0* True negative=7* Sensitivity=100 (95% CI: 62.9-100)* Specificity=93.3(95% CI: 63.7-93.3)* Positive LR= 15.0 (95% CI: 1.73-15.0)* Negative LR= 0 (95% CI: 0-0.58)* AUROC=1.00 (95% CI: 1.00-1.00) Forns ≥ 6.8 versus recent guidelines for detection of oesophageal varices for all patients n=23 True positive=6* False positive=2* False negative=0* True negative=15* Sensitivity=100 (95% CI: 58.9-100)*	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Specificity=88.2 (95% CI: 73.7-88.2)* Positive LR= 8.5 (95% CI: 2.2-8.5)* Negative LR= 0 (95% CI: 0-0.56)* AUROC=0.98 (95% CI: 0.93-1.00) Forns ≥ 6.8 versus recent guidelines for detection of oesophageal varices for CFLD patients n=13 True positive=6* False positive=1* False negative=0* True negative=6* Sensitivity=100 (95% CI: 62.9-100)* Specificity=85.7 (95% CI: 53.9-85.7)* Positive LR= 7.0 (95% CI: 1.37-7.0)* Negative LR= 0 (95% CI: 0-0.69)* AUROC=0.98 (95% CI: 0.91-1.00) *Calculated by the NGA technical team from data reported in the article **0.5 person was added by the NGA technical	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				team to calculate likelihood ratios and 95% confidence intervals LR = likelihood ratio	
Full citation Lewindon, P. J., Shepherd, R. W., Walsh, M. J., Greer, R. M., Williamson, R., Pereira, T. N., Frawley, K., Bell, S. C., Smith, J. L., Ramm, G. A., Importance of hepatic fibrosis in cystic fibrosis and the predictive value of liver biopsy, Hepatology, 53, 193-201, 2011 Ref Id 332925 Country/ies where the study was carried out Australia Study type Prospective cohort study Aim of the study To evaluate dual-pass liver biopsy and the commonly used clinical tools	Sample size 40 children with suspected cystic fibrosis liver disease Characteristics 24 females 16 males Age: 2.38-18.73 years at enrollment Median age=10.64 years 96% Caucasian 20% had cystic fibrosis related diabetes 68% f508 homozygotes Median FEV1 value=83.5% 9/40 had portal hypertension (PHT) Inclusion Criteria Patients with suspected cystic fibrosis defined as the following:	Tests Reference standard -Dual pass percutaneous liver biopsy with US guidance under general anesthesia (14-Fr Tru-Cut, throw length =20 mm) from the right lobe via the same skin incision with different angles of insertion. The tissue was immediately fixed in 10% buffered formalin and embedded in paraffin. Liver sections were evaluated by a hepatopathologist (Richard Williamson) blinded to the clinical data; more than 10 levels of tissue sections stained with hematoxylin and eosin or hematoxylin and Van Gieson's stain were used. For fibrosis scoring, the Scheuer F0-F4 staging system was used (F0	Methods Setting: major cystic fibrosis referral clinic of the Royal Children's Hospital (Brisbane, Australia) At enrollment, the following were performed or determined for all patients: history, physical examination, Df508 genotype, lung function, serum amino-transferases, liver synthetic function (international normalized ratio and albumin), and liver US as well as upper gastrointestinal endoscopy, serum draw for research, and dual-pass liver biopsy under general anesthesia. Follow-up were up to 12 years, until death, transplantation, or	Results n=40 patients Ultrasound versus biopsy True positive=25* False positive=5* False negative=6* True negative=4* Sensitivity=0.81 (95% CI: 0.73-0.89)* Specificity=0.44 (95% CI: 0.17-0.73)* Positive LR= 1.45 (95% CI: 0.87-3.3)* Negative LR= 0.44 (95% CI: 0.15-1.64)* AUROC=0.63 (95% CI: NR) Clinical exam-Hepatomegaly (HM) versus biopsy True positive=21* False positive=6* False negative=10* True negative=3* Sensitivity= 0.68 (95% CI: 0.61-0.77)*	Limitations QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? N/A 2.A Could the conduct or interpretation of the index test have introduced bias? LOW RISK OF BIAS

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
available to clinicians when they are confronted with a patient with suspected CFLD. To look at the ability of the latter to predict hepatobiliary fibrosis on biopsy, and compare the value of biopsy to the value of clinical modalities currently used to predict adverse outcomes (i.e.,PHT and/or liver failure) and mortality over prolonged clinical follow-up (up to 12 years). Study dates Between 1999 and 2009 Source of funding National Health and Medical Research Council of Australia and Royal Children's Hospital Foundation of Brisbane	-hepatomegaly (HM) with or without splenomegaly -a persistent (>6-month) elevation of serum alanine aminotransferase (ALT; level > 1.5 x upper limit of normal) -abnormal liver US findings (abnormal echogenicity or a nodular edge) Exclusion Criteria Patients with liver synthetic dysfunction or a history of hepatobiliary surgery	=no fibrosis, F4 = cirrhosis). Only sections with at least five portal tracts were deemed adequate for assessment. Index tests -Clinical examinations: Hepatomegaly with or without splenomegaly -Serum ALT levels performed at enrollment -Ultrasound images were obtained after fasting to induce gallbladder distension, using real-time scanners: Acuson Sequoia (Siemens Medical, Erlangen, Germany) with 2.5- to 4-MHz or 5.5- to 8.5-MHz probes or ATL HDI 5000 (Philips Medical Systems, Best, the Netherlands) with 2- to 5-MHz or 5- to 7-MHz probes. Sonographic images were reviewed by a pediatric radiologist (Kieran Frawley) blinded to	survival as of March 2009. All patients received standard CF pulmonary and nutritional care, all patients with biopsy-confirmed fibrosis were prescribed ursodeoxycholic acid (15 mg/kg/day), all patients were reviewed at least on a 6-month basis. For the purposes of this study, prospectively recorded follow-up data included clinical progress, occurrence of cystic fibrosis–related diabetes mellitus (CFRD; defined as insulin-dependent diabetes mellitus), survival, solid organ transplantation, forced expiratory volume in 1 second (FEV1), liver aminotransferases, liver synthetic function, and occurrence of PHT.	Specificity= 0.33 (95% CI: 0.10-0.65)* Positive LR=1.02 (95% CI: 0.67-2.23)* Negative LR=0.97 (95% CI: 0.35-4.11)* AUROC=0.51 (95% CI: NR) ALT versus biopsy True positive=0.5** False positive=0.5** False negative=17* True negative=23* Sensitivity=0.03 (95% CI: 0-0.06)* Specificity=0.98 (95% CI: 0.96-1.0)* Positive LR=1.34 (95% CI: 0-1408086.43)* Negative LR=0.99 (95% CI: 0.94-1.04)* AUROC=0.59 (95% CI: NR) US+HSM+LFT versus biopsy for F1-F4 fibrosis HSM=hepatosplenomegaly True positive=17* False positive=20* False negative=0.5** True negative=3*	2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN Reference Standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3.A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3.B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW RISK Flow and Timing Was there an appropriate interval between index test(s) and reference standard? No, both tests conducted at enrollment Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		clinical and biopsy findings and previous interpretations. Briefly, liver images were recorded as nodular edge, nodular, heterogeneous, or normal echogenicity with or without splenomegaly. Normal US was defined as normal echogenicity with no splenomegaly. US evidence of PHT included a nodular liver with splenomegaly.		Sensitivity=0.97 (95% CI: 0.85-1.0)* Specificity=0.13 (95% CI: 0.04-0.15)* Positive LR=1.12 (95% CI: 0.89-1.18)* Negative LR=0.22 (95% CI: 0-3.6)* AUROC=0.69 (95% CI: NR) US+HSM+LFT versus biopsy for F2-F4 significant fibrosis True positive=14* False positive=12* False negative=3* True negative=11* Sensitivity=0.82 (95% CI: 0.62-0.95)* Specificity=0.48 (95% CI: 0.33-0.57)* Positive LR=1.58 (95% CI: 0.93-2.22)* Negative LR=0.37 (95% CI: 0.09-1.15)* AUROC=0.68 (95% CI: NR) LR = likelihood ratio NC=not calculable NR=not reported	Were all patients included in the analysis? Yes 4.A Could the patient flow have introduced bias? LOW RISK Other information

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				*Calculated by the NGA technical team from data reported in the article **0.5 person was added by the NGA technical team to calculate likelihood ratios and 95% confidence intervals.	
Full citation Lindblad, A., Glaumann, H., Strandvik, B., Natural history of liver disease in cystic fibrosis, Hepatology, 30, 1151-8, 1999 Ref Id 329857 Country/ies where the study was carried out Sweden Study type Prospective cohort study Aim of the study To evaluate the natural history of CF-associated liver disease over a 15-year period in a well-controlled population of	Sample size n=124 followed up during 1976-1993 n=27 received biopsy in 1976-1979 n=41 received biopsy in 1989-1993 Characteristics 41 patients who received biopsy in 1989-1993 Median age 19 years, range 5 to 43 years Further characteristics details on 41 patients not reported in the study Clinical Data on All Patients With CF Attending the	Tests Reference standard Liver biopsy performed under general anesthesia in patients younger than 16 years and under local anesthesia in older patients. The biopsy specimen was prepared according to routine methods and stained with hematoxylineosin, periodic acid-Schiff diastase treatment, reticulin, and iron stains. The biopsies were evaluated regarding fibrosis (normal; slight, enlarged portal zones; moderate, tendency towards septa formation; severe, bridging fibrosis; and cirrhosis, complete	Methods Setting: Stockholm Cystic Fibrosis Center All patients had pathological sweat tests (chloride .60 mmol/L). Patients with pancreatic insufficiency were treated with pancreatic enzymes (enteric-coated microspheres after 1982) and multivitamins including vitamin A. During the entire study, additional vitamin E in water-soluble form was prescribed to all patients, as was the oral mucolytic bromhexine and inhalation of salbutamol and saline and/or N-acetyl cysteine. Patients chronically colonized with Pseudomonas	Results Results are only for 1989-1993 n=41 AUROC not reported for all tests For moderate or severe fibrosis and cirrhosis outcome LFT versus biopsy True positive=14* False positive=15* False negative=0* True negative=12* Sensitivity=1.0 (95% CI: 0.78-1.0)* Specificity=0.44 (95% CI: 0.33-0.44)* Positive LR= 1.8 (95% CI: 1.17-1.8)* Negative LR= 0 (95% CI: 0-0.67)* Ultrasound versus biopsy	Limitations QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? N/A

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
patients with CF by biochemical markers, liver biopsies, and, during the latest years, also ultrasonography (US). Study dates 1976-1993 Source of funding Swedish Medical Research Council	CF Center During the Year 1993 n=140 Women/men: 75/65 Age, year-mean (median): 16.5 (15) BMI mean (median): 21 (21) 45 homozygous for F508 genotype 40 heterozygous for F508 genotype Mean FEV 1.0, %: 70 Number of patients with late diagnosis (older than 10 years of age), n (%): 13 (9) Inclusion Criteria All patients with Cystic Fibrosis cared for at the Stockholm CF center and attended the center 2 or more times between 1976 and 1993 Exclusion Criteria -Patients who were seen only once during the	septa with regenerative noduli). Steatosis, bile duct proliferation, and inflammation were classified as absent, slight, moderate, or severe. A minimum of 4 portal zones were evaluated in each biopsy. Index tests - Ultrasonography characterized as normal or pathological (increased and/or irregular echogenicity) -Liver function test included serum activities of alanine transaminase (ALT), aspartate transaminase (AST), and g-glutamyltransferase (gGT) (with an upper reference level of ,0.8, ,0.8, and ,0.5 µkata/l, respectively). -Combined US and LFT	aeruginosa, Stenotrophomonas maltophilia, and Burkholderia cepacia were treated with intravenous antibiotics (an aminoglycoside and a b-lactam) for minor signs of exacerbations, whereas patients not colonized with these organisms were treated with oral antibiotics covering Staphylococcus aureus and/or Haemophilus influenzae, generally flucloxacillin or trimethoprim-sulpha. After 1985, most intravenous antibiotic courses were given at home. Intralipid 10% (Kabi, Stockholm, Sweden) at a dose of 10 mL/kg body weight was given regularly to most patients in connection with intravenous courses of antibiotics or at signs of failure to thrive. All patients were investigated annually for inflammatory status and	True positive=12* False positive=8* False negative=2* True negative=19* Sensitivity=0.86 (95% CI: 0.61-0.97)* Specificity=0.70 (95% CI: 0.58-0.76)* Positive LR= 2.9 (95% CI: 1.45-4.13)* Negative LR= 0.2 (95% CI: 0.03-0.67)* US+LFT versus biopsy True positive=12* False positive=7* False negative=2* True negative=20* Sensitivity=0.86 (95% CI: 0.62-0.97)* Specificity=0.74 (95% CI: 0.62-0.80)* Positive LR= 3.31 (95% CI: 1.6-4.9)* Negative LR= 0.19 (95% CI: 0.03-0.63)* For moderate or severe fibrosis and cirrhosis and/or moderate to severe steatosis outcome LFT versus biopsy True positive=19* False positive=10*	2.A Could the conduct or interpretation of the index test have introduced bias? UNCLEAR 2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN Reference Standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Unclear 3.A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK-there were 10 people who had no biopsy but were identified to have biochemical liver disease 3.B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW RISK Flow and Timing Was there an appropriate interval between index test(s) and reference standard? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
	period at the CF center -Chronic hepatitis B infection and late diagnosis of CF at the age over 60 -Less than 4 years of age		lung function tests including forced vital capacity (FVC), forced expiratory volume in one second (FEV 1.0), chest radiograph, serum levels of retinol and a-tocopherol, liver function tests (LFTs), and the fatty acid pattern of serum phospholipids. After 1989, ultrasonography (US) of the liver was also performed annually. Antipyrine and galactose elimination capacity tests were performed in connection with liver biopsies. During the period 1976 to 1979, percutaneous liver biopsy was performed in 27 patients (median age 11 years, range 2 to 27 years), in most patients it was performed at least twice with a 1- to 3-year interval. During the years 1980 to 1988 very few biopsies were performed. From 1989 to 1993, liver	False negative=4* True negative=8* Sensitivity=0.83 (95% CI: 0.68-0.94)* Specificity=0.44 (95% CI: 0.26-0.58)* Positive LR= 1.49 (95% CI: 0.92-2.25)* Negative LR= 0.39 (95% CI: 0.11-1.22)* Ultrasound versus biopsy True positive=16* False positive=4* False negative=7* True negative=14* Sensitivity=0.70 (95% CI: 0.54-0.80)* Specificity=0.78 (95% CI: 0.58-0.92)* Positive LR= 3.13 (95% CI: 1.3-9.5)* Negative LR= 0.39 (95% CI: 0.22-0.8)* US+LFT versus biopsy True positive=15* False positive=4* False negative=8* True negative=14* Sensitivity=0.65 (95% CI: 0.5-0.76)* Specificity=0.78 (95% CI: 0.58-0.92)*	Did all patients receive a reference standard? No, 10 out of 41 patients did not receive biopsy Did patients receive the same reference standard? Yes Were all patients included in the analysis? No-only 41 out of 124 were analysed 4.A Could the patient flow have introduced bias? HIGH RISK Other information

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
			biopsies were performed in 41 patients, aged 5 years or older (median age 19 years, range 5 to 43 years). During the last part of the study 13 patients were prescribed ursodeoxycholic acid (UDCA) and 12 of them were followed up for 2 years with biopsies. Only the first biopsy before treatment with UDCA was evaluated in the present study. Definitions: Biochemical liver disease (BLD) was defined as elevation above the upper reference level of any serum liver enzyme included in the LFT for at least 2 consecutive years in patients 4 years of age or older. A patient was thereafter classified as BLD even if LFT results were later normalized. Clinical liver disease was defined as multilobular cirrhosis (MLC) and	Positive LR= 2.94 (95% CI: 1.18-9.1)* Negative LR= 0.45 (95% CI: 0.26-0.87)* LR = likelihood ratio *Calculated by the NGA technical team from data reported in the article	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
			always included clinical (hepato) splenomegaly with esophageal varices or signs of hypersplenism and biopsy-proven cirrhosis. All other patients were classified as having no liver disease (NLD). Statistical Analysis: An analysis of variance test, the Mann-Whitney U test, the x2 test with Yates'correction, Fisher's exact test, and the Kruskal-Wallis test were used when appropriate. The level of significance was set to 0.05.		
Full citation Mueller-Abt, P. R., Frawley, K. J., Greer, R. M., Lewindon, P. J., Comparison of ultrasound and biopsy findings in children with cystic fibrosis related liver disease, Journal of Cystic Fibrosis, 7, 215-21, 2008 Ref Id	Sample size n=30 children with CF Characteristics 13 girls/17 boys Mean age, years: 10 Age range: 11 months to 17 years Inclusion Criteria	Tests Reference standard Percutaneous liver biopsy using ultrasound guidance. The ultrasound was used for biopsy guidance only and a detailed ultrasound assessment of the liver was not performed at the time of the biopsy. Two samples, to limit sampling error, were	Methods Setting: CF clinic A retrospective analysis of ultrasound findings was per- formed in 30 CF-patients (13 girls, 17 boys) who underwent a liver biopsy and ultrasound between April 1997 and September 2003. The CF-patients undergoing liver biopsy were	Results n=30 Ultrasound versus biopsy for liver disease or cirrhosis outcome True positive=15* False positive=3* False negative=8* True negative=4* Sensitivity=0.65 (95% CI: 0.55-0.74)* Specificity=0.57 (95% CI: 0.22-0.87)*	Limitations QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
354053 Country/ies where the study was carried out Australia Study type Retrospective cohort study Aim of the study To determine if hepatic ultrasound findings in paediatric patients with cystic fibrosis and suspected liver disease are related to histopathological results derived from liver biopsies. Study dates 1997-2003 Source of funding Not reported	-All patients attending the CF clinic with positive sweat tests confirming CF. -Evidence of biochemical liver disease (persistent elevation of ALT above upper limit of normal over a period of at least 6 months) -Clinical hepatomegaly or hepatosplenomegaly -Sonographic evidence of liver disease Exclusion Criteria Not reported	obtained from the right lobe using a triggered trucut to obtain 20 mm cores. In all specimens, at least 6 portal tracts were available for analysis and a Scheuer grading for fibrosis was allocated to each patient by a histopathologist blinded to US findings. Scheuer-Score of 0 was regarded as normal, a score of 1–2 as mild to moderate reversible periportal changes and 3–4 was assessed as definite fibrosis/cirrhosis. Index standard US scans were obtained after a 4-hour fast in children under 2 years and a 6- hour fast in children over 2 years for gallbladder distension. Sonographic images were independently reviewed two times on hardcopies by a pediatric radiology fellow (investigator 1) and an experienced	identified from the cystic fibrosis clinic database. Ethical approval of this study was granted by the institutional ethics committee as part of a wide study into liver fibrosis in cystic fibrosis. All patients were attending the CF clinic with positive sweat tests confirming CF. Patients underwent liver biopsy if two out of three of the following criteria were fulfilled: 1. evidence of biochemical liver disease (persistent elevation of ALT above upper limit of normal over a period of at least 6 months), 2. clinical hepatomegaly or hepatosplenomegaly, 3. sonographic evidence of liver disease. Informed consent was obtained from the parents for the biopsy. The time interval between biopsy and ultrasound was between 0 and 183 days (mean 42 days).	Positive LR= 1.52 (95% CI: 0.7-5.78)* Negative LR= 0.61 (95% CI: 0.29-2.06)* AUROC NR Ultrasound versus biopsy for cirrhosis outcome only True positive=8* False positive=1* False negative=6* True negative=15* Sensitivity=0.57 (95% CI: 0.36-0.64)* Specificity=0.94 (95% CI: 0.75-1.00)* Positive LR= 9.14 (95% CI: 1.47-192.8)* Negative LR= 0.46 (95% CI: 0.36-0.85)* AUROC NR NR= not reported LR = likelihood ratio *Calculated by the NGA technical team from data reported in the article	1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? N/A 2.A Could the conduct or interpretation of the index test have introduced bias? LOW RISK OF BIAS 2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN Reference Standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3.A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3.B Is there concern that the target condition as defined by

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		paediatric radiologist (investigator 2). The reviewers were unaware of the clinical findings and previous interpretation and were blinded to the histology. After independent review a consensus result was reached in cases with differing interpretations for each of the ultrasound criteria evaluated. A summary interpretation of the findings was performed by each reviewer. There were three categories: normal, indeterminate (suggestion of liver disease but no definite signs of cirrhosis) and cirrhosis. Cases without liver abnormality were graded as normal. Increased hepatic echogenicity, heterogeneity and/or increased attenuation in the absence of nodularity of the liver surface were classified as indeterminate. Splenomegaly as an isolated finding was			the reference standard does not match the review question? LOW CONCERN Flow and Timing Was there an appropriate interval between index test(s) and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4.A Could the patient flow have introduced bias? LOW RISK Other information

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		also regarded as indeterminate. All patients with nodularity of the liver surface were classified as cirrhosis.			
Full citation Patriquin, H., Lenaerts, C., Smith, L., Perreault, G., Grignon, A., Filiatrault, D., Boisvert, J., Roy, C. C., Rasquin-Weber, A., Liver disease in children with cystic fibrosis: US-biochemical comparison in 195 patients, Radiology, 211, 229-32, 1999 Ref Id 333103 Country/ies where the study was carried out Canada Study type Prospective cohort study Aim of the study To determine if abnormal liver architecture at ultrasonography	Sample size n=195 children Characteristics 112 boys, 83 girls; mean age, 8.5 years, age range 1-23 years Inclusion Criteria Children with CF Exclusion Criteria Not reported	Tests Reference test US: US scans were obtained without sedation after a 4-hour fast in children aged 2–6 years and after an 8-hour fast in patients older than 6 years. One of the following commercially available machines was used: Ultramark 5, 8, or 9 (Advanced Technology Laboratories, Seattle, Wash) or Quantum II (Siemens Medical Systems, Erlangen, Germany) with a 3.5-, 5.0-, or 7.0-MHz transducer. The sonograms were obtained by one of five pediatric radiologists (H.P., G.P., A.G., D.F., J.B.) and were later reviewed by one of the five. No radiologist was aware of the	Methods Setting: CF clinic For 1 year, all 195 children (112 boys, 83 girls; mean age, 8.5 years) attending a CF clinic underwent abdominal US and a standard set of liver function tests. Aspartate aminotransferase, alanine aminotransferase, and □-glutamyltransferase levels were analyzed. US signs were interpreted as follows: hypoechogenicity with prominent portal tracks as edema, hyperechogenicity as steatosis, and increased attenuation and nodules within or at the edge of the liver as cirrhosis. Signs of portal hypertension also were sought. US signs were compared with liver function test results.	Results LFT: ALT versus US n=195 True positive=24* False positive=33* False negative=14* True negative=124* Sensitivity=63.2 (95% CI: 48.0-76.3)* Specificity=79.0 (95% CI: 75.3-82.2)* Positive LR= 3.0 (95% CI: 1.95-4.28)* Negative LR= 0.47 (95% CI: 0.29-0.69)* AUROC=NR GGT versus US n=195 True positive=19* False positive=15* False negative=19* True negative=142* Sensitivity=50.0 (95% CI: 36.2-62.4)* Specificity=90.4 (95% CI: 87.1-93.4)*	Limitations QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? N/A 2.A Could the conduct or interpretation of the index test

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
(US) is related to abnormal function in children with cystic fibrosis (CF). Study dates 1999 Source of funding Not reported		biochemical results at the time of examination or review. US included a survey of the entire abdomen as well as a detailed examination of liver architecture. This included liver echogenicity, which was compared with that of the renal cortex. The liver was called hyperechoic if it was brighter than the cortex of the right kidney and if the walls of portal veins were difficult to distinguish from the adjacent liver parenchyma. Sound attenuation by the liver was assessed and was considered to be increased if the posterior surface of the liver was not visible with a transducer frequency that allowed sound penetration and depiction of the kidney through the liver. Evidence of nodules within and at the surface of the liver was sought. The		Positive LR= 5.23 (95% CI: 2.80-9.53)* Negative LR= 0.55 (95% CI: 0.40-0.73)* AUROC=NR AST versus US n=195 True positive=18* False positive=19* False negative=20* True negative=138* Sensitivity=47.4 (95% CI: 33.4-60.6)* Specificity=87.9 (95% CI: 84.5-91.1)* Positive LR= 3.91 (95% CI: 2.16-6.80)* Negative LR= 0.60 (95% CI: 0.43-0.79)* AUROC=NR *Calculated by the NGA technical team from data reported in the article NR=not reported LR = likelihood ratio	have introduced bias? LOW RISK 2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN Reference Standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3.A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3.B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW RISK Flow and Timing Was there an appropriate interval between index test(s) and reference standard? No, US and LFTs performed on same day Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		caliber of the intrahepatic bile ducts was noted, and they were termed dilated if they exceeded 2 mm in diameter. Evidence of portal hypertension was sought (spleno-megaly, collateral veins, lesser omental thick- ening); when found, Doppler US was performed. The presence and direction of blood flow in the splanchnic and intrahepatic portal veins was assessed, and portosystemic collateral vessels, especially esophageal varices, were sought. US abnormalities of liver architecture were interpreted as follows: hyperechogenicity as steatosis and heteroechogenicity of liver architecture accompanied by increased sound attenuation as cirrhosis. Nodules within or at the edge of the liver were also interpreted as cirrhosis. Hypoechoic liver parenchyma and			Were all patients included in the analysis? Yes 4.A Could the patient flow have introduced bias? LOW RISK Other information

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		bright periportal echoes of normal thickness also were noted, but no pathologic interpretation was attributed to these findings. Index tests Liver function tests included total and direct bilirubin, aspartate aminotransferase (AST), alanine aminotransferase (ALT), □glutamyltransferase (GGT), albumin, pre-albumin, prothrombin time, and fasting and postprandial endogenous bile acid (cholylglycine) tests.			
Full citation Rath, T., Hage, L., Kugler, M., Menendez Menendez, K., Zachoval, R., Naehrlich, L., Schulz, R., Roderfeld, M., Roeb, E., Serum Proteome Profiling	Sample size n=45 Characteristics CFLD n=17/ 53% male No CFLD n=28/ 61% male Mean age, y (SD): no CFLD- 21.4	Tests Reference test Diagnosis of CFLD was established according to recent guidelines if least two of the following conditions on at least two consecutive examinations	Methods 45 CF patients were included in the study and received transient elastography. Differential regulation of 220 different serum proteins was assessed in a subgroup of patients with and without CFLD. Most	Results n=45 APRI versus recent guidelines for detection of CFLD True positive=8* False positive=2* False negative=9*	Limitations QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Unclear Was a case-control design avoided? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Identifies Novel and Powerful Markers of Cystic Fibrosis Liver Disease, PLoS ONE, 8, 2013 Ref Id 340488 Country/ies where the study was carried out Germany Study type Prospective cohort study Aim of the study To identify new experimental biomarkers for the detection of CFLD. Study dates 2008-2010 Source of funding Deutsche Forschungsgemeinschaft (RO 957/7-1 and RO 957/8-1), a research Grant of the University Medical Center Giessen and Marburg (UKGM 10/2010 GI), and from ZooMAP (01KI1003E, Bundesministerium	(11.8); CFLD-29 (10.8) Inclusion Criteria -Diagnosis of CF was established by sweat test and later confirmed by genetic tests in all subjects Exclusion Criteria -Other causes for chronic liver disease	spanning a one-year period were present: (i) Hepatomegaly (liver span >2 cm below the costal margin on the medioclavicular line) confirmed by ultrasound, (ii) two abnormal serum liver enzyme levels (ALT, AST, γGT > ULN), (iii) ultrasound abnormalities other than hepatomegaly (increased, heterogeneous echogenicity, nodularity, irregular margins). Index tests -Transient elastography (TE): Liver stiffness by TE was evaluated using the same FibroScan® (Echosens, Paris, France) device in all patients. Non-invasive measurements were performed by a single experienced investigator blinded to the clinical status of the patients on the right lobe of the liver	interesting candidate proteins were further quantified and validated by ELISA in the whole patient cohort. To assess a potential relation of biomarker expression to the degree of hepatic fibrosis, serum biomarkers were further determined in 18 HCV patients where liver histology was available.	True negative=27* Sensitivity=47.1 (95% CI: 28.2-56.7)* Specificity=93.1 (95% CI: 82.0-98.7)* Positive LR= 6.82 (95% CI: 1.57-44.7)* Negative LR= 0.57 (95% CI: 0.44-0.88)* AUROC=0.75 (95% CI: 0.58-0.91) NCC estimates based upon information in the paper--n adds up to 46 due to rounding errors (i.e. they haven't given sensitivities to a great enough degree of accuracy). ALP versus recent guidelines for detection of CFLD True positive=12* False positive=5* False negative=5* True negative=23* Sensitivity=70.6 (95% CI: 49.5-85.5)* Specificity=82.1 (95% CI: 69.3-91.2)* Positive LR= 3.95 (95% CI: 1.61-9.74)* Negative LR= 0.36 (95% CI: 0.16-0.73)*	Did the study avoid inappropriate exclusions? Yes 1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? N/A 2.A Could the conduct or interpretation of the index test have introduced bias? LOW RISK OF BIAS 2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN Reference Standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
fÃ¼r Bildung und Forschung, BMBF)		through the intercostal space at a depth of 25 and 65 mm from skin surface. In children below 15 kg of weight the FibroScan® S probe, developed for liver stiffness measurements in children, was used. For each patient, the stiffness value was calculated as the median of ten successful measurements. TE was considered valid if 10 successful measurements with a success rate ≥ 60% and an interquartile range ≤ 30% of the median were obtained. Results are expressed in kilopascal (kPa). Total examination time was approximately 5 minutes per patient. -Alkaline phosphatase (ALP) -AST/Platelets-Ratio-Index (APRI)		AUROC=0.61 (95% CI: 0.44-0.79) TE versus recent guidelines for detection of CFLD True positive=14* False positive=0.5** False negative=3* True negative=28* Sensitivity=82.4 (95% CI: 64.2-85.3)* Specificity=98.2 (95% CI: 87.4-100)* Positive LR= 46.9 (95% CI: 5.1-25489647)* Negative LR= 0.18 (95% CI: 0.15-0.41)* AUROC=0.91 (95% CI: 0.78-1.00) LR = likelihood ratio NC=not calculable NR=not reported *Calculated by the NGA technical team from data reported in the article **0.5 person was added by the NGA technical team to calculate likelihood ratios and 95% confidence intervals.	3.A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3.B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW RISK Flow and Timing Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4.A Could the patient flow have introduced bias? LOW RISK Other information
Full citation	Sample size	Tests	Methods	Results	Limitations

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Rath, T., Menendez, K. M., Kugler, M., Hage, L., Wenzel, C., Schulz, R., Graf, J., Nahrlich, L., Roeb, E., Roderfeld, M., TIMP-1/-2 and transient elastography allow non invasive diagnosis of cystic fibrosis associated liver disease, Digestive & Liver Disease, 44, 780-7, 2012 Ref Id 354071 Country/ies where the study was carried out Germany Study type Prospective cohort study Aim of the study Compare the value of transient elastography and experimental fibrosis markers for the detection of liver disease in CF patients Study dates	145 CF patients (75 children, 70 adults) Characteristics Paediatric CF patients No CFLD (n=45)/CFLD (n=30) Male, %: 60/30 Age, years, mean (SD): 10.9 (4.9)/10.6 (4.3) Adult CF patients No CFLD (n=32)/CFLD (n=29)/CFLD + PHT Male, %: 53/48/66 Age, years, mean (SD): 32.3 (9.3)/30.6 (8.6)/32.2 (5.8) Inclusion Criteria Diagnosis of CF was established by sweat test and later confirmed by genetic tests in all subjects Exclusion Criteria Other causes for chronic LD	Reference test Diagnosis of CFLD was established according to recent guidelines if least two of the following conditions on at least two consecutive examinations spanning a one-year period were present: (i) Hepatomegaly (liver span >2 cm below the costal margin on the medioclavicular line) confirmed by ultrasound, (ii) two abnormal serum liver enzyme levels (ALT, AST, γGT > ULN), (iii) ultrasound abnormalities other than hepatomegaly (increased, heterogeneous echogenicity, nodularity, irregular margins). Diagnosis of PHT was based on clinical and lab data combined with sonographic or endoscopic signs of PHT. Index tests	145 CF patients (75 children, 70 adults) were prospectively studied and received transient elastography. CF liver disease was diagnosed according to recent guidelines. Serum concentrations of YKL-40, HA, PIIIP, MMP-9, TIMP-1 and TIMP-2 were determined by ELISA.	TE at 5.5 kPA cut-off versus recent guidelines for detection of CFLD only-all CF patients n=136 True positive=39* False positive=11* False negative=35* True negative=51* Sensitivity=52.7 (95% CI: 44.9-58.9)* Specificity=82.3 (95% CI: 72.9-89.7)* Positive LR= 2.97 (95% CI: 1.65-5.70)* Negative LR= 0.58 (95% CI: 0.46-0.76)* AUROC=0.68 (95% CI: 0.59-0.77) TE at 5.5 kPA cut-off versus recent guidelines for detection of CFLD only-adult CF patients n=61 True positive=16* False positive=7* False negative=13* True negative=25* Sensitivity=55.2 (95% CI: 40.7-66.8)* Specificity=78.1 (95% CI: 65.0-88.7)*	QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? N/A 2.A Could the conduct or interpretation of the index test have introduced bias? LOW RISK OF BIAS 2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
2008-2010 Source of funding Deutsche Forschungsgemeinschaft, ZooMAP, University Medical Ctr Giessen and Marburg		-Transient elastography (TE): Liver stiffness by TE was evaluated using the same FibroScan® (Echosens, Paris, France) device in all patients. Non-invasive measurements were performed by a single experienced investigator blinded to the clinical status of the patients on the right lobe of the liver through the intercostal space at a depth of 25 and 65 mm from skin surface. In children below 15 kg of weight the FibroScan® S probe, developed for liver stiffness measurements in children, was used. For each patient, the stiffness value was calculated as the median of ten successful measurements. TE was considered valid if 10 successful measurements with a success rate ≥ 60% and an interquartile range ≤ 30% of the		Positive LR= 2.52 (95% CI: 1.16-5.89)* Negative LR= 0.57 (95% CI: 0.38-0.91)* AUROC=0.69 (95% CI: 0.56-0.81) TE at 5.5 kPA cut-off versus recent guidelines for detection of CFLD only-paediatric CF patients n=75 True positive=24* False positive=7* False negative=21* True negative=23* Sensitivity=53.3 (95% CI: 43.2-61.2)* Specificity=76.7 (95% CI: 61.4-88.4)* Positive LR= 2.29 (95% CI: 1.12-5.28)* Negative LR= 0.61 (95% CI: 0.44-0.93)* AUROC=0.68 (95% CI: 0.56-0.81) TE at 11.5 kPA cut-off versus recent guidelines for detection of CFLD and PHT-adult CF patients n=70 True positive=6*	Reference Standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3.A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3.B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW RISK Flow and Timing Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4.A Could the patient flow have introduced bias? LOW RISK Other information

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		median were obtained. Results are expressed in kilopascal (kPa). Total examination time was approximately 5 minutes per patient.		False positive=1* False negative=3* True negative=60* Sensitivity=66.7 (95% CI: 36.2-77.2)* Specificity=98.4 (95% CI: 93.9-99.9)* Positive LR= 40.67 (95% CI: 5.91-877.4)* Negative LR= 0.34 (95% CI: 0.23-0.68)* AUROC=0.86 (95% CI: 0.66-1.00) A cut-off of 5.5 kPa was optimal for TE for the diagnosis of CFLD in every patient cohort, whereas a cut-off of 11.5 kPa was optimal for TE for the diagnosis of PHT in adult CF patients with PHT LR = likelihood ratio *Calculated by the NGA technical team from data reported in the article	
Full citation Sadler, M. D., Crotty, P., Fatovich, L., Wilson, S., Rabin, H. R., Myers, R. P., Noninvasive	Sample size n=127 Characteristics All patients n=127 Age, median years	Tests Reference test Diagnosis of CFLD was established according to previously published criteria if least two of	Methods Setting: Adult CF clinic of Calgary and Southern Alberta At enrollment, patient demographics, anthropometric	Results LSM using TE versus published criteria for CFLD diagnosis n=127 ≥3.7 kPa	Limitations QUADAS 2 checklist Patient selection

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
methods, including transient elastography, for the detection of liver disease in adults with cystic fibrosis, Canadian Journal of Gastroenterology & Hepatology, 29, 139-44, 2015 Ref Id 354082 Country/ies where the study was carried out Canada Study type Prospective cohort study Aim of the study To evaluate the diagnostic performance of noninvasive methods for the detection of CFLD with a focus on transient elastography (TE) Study dates 2010-2011 Source of funding AI-HS, CIHR, and Canadian Liver Foundation	(interquartile range): 27 (22-37) Male (%): 60 (47) 25% were prescribed UDCA With CFLD n=18 Age, median years (interquartile range): 28 (18-32) Male (%): 10 (56) 83% were prescribed UDCA Without CFLD n=109 Age, median years (interquartile range): 27 (22-37) Male (%): 50 (42) 14% were prescribed UDCA Inclusion Criteria ≥18 years of age with CF Exclusion Criteria Hepatitis B or C	the following conditions were present: (i) Hepatomegaly and/or splenomegaly confirmed by ultrasonography, (ii) abnormal liver biochemistry consisting of elevated levels of any two of ALT, AST, or GGT, (iii) ultrasound abnormalities other than hepatomegaly (increased, heterogeneous echogenicity, nodularity, irregular margins, splenomegaly presence). Index tests -Liver stiffness measurement by transient elastography (TE) using FibroScan® probe. -Aspartate aminotransferase to Platelets-Ratio-Index (APRI) was calculated as (AST/upper limit of normal for AST) x	measurements, CF transmembrane regulator genetic mutations, UDCA use and history of CF-related complications, diabetes mellitus and lung transplantation were recorded. All patients underwent a physical exam and routine lab investigations. Individuals with examination findings suggestive of liver disease or abnormal liver biochemistry underwent ab ultrasonography (n=78). Spirometry values from pulmonary function testing on the day of enrollment were also recorded. Patients at the Adult CF Clinic of Calgary and Southern Alberta (n=127) underwent liver stiffness measurement (LSM) by TE using the FibroScan (FS, Ecosens, France) M probe; aspartate aminotransferase to platelet ratio index (APRI) and FibroTest (FT) scores were also	True positive=16* False positive=69* False negative=2* True negative=40* Sensitivity=89 (95% CI: 66-98)* Specificity=37 (95% CI: 33-38)* Positive LR= 1.40 (95% CI: 0.98-1.59)* Negative LR= 0.30 (95% CI: 0.05-1.04)* AUROC NR ≥5.3 kPa** True positive=12* False positive=19* False negative=6* True negative=90* Sensitivity=67 (95% CI: 43-85)* Specificity=83 (95% CI: 79-86)* Positive LR= 3.83 (95% CI: 2.04-5.87)* Negative LR= 0.40 (95% CI: 0.18-0.72)* AUROC=0.78 (95% CI: 0.65-0.92) >6.0 kPa True positive=10* False positive=10*	Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? N/A 2.A Could the conduct or interpretation of the index test have introduced bias? UNCLEAR 2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN Reference Standard

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		(100/platelets (x109/L)). -FibroTest (FT) was calculated based on age, sex, GGT, total bili-alpha-2macroglobulin, apolipoproten A1 and haptoglobin.	calculated. The diagnostic performance of these tools for the detection of CFLD (defined as two or more the following criteria: abnormal liver biochemistry, hepatomegaly or sonographic abnormalities other than steatosis) were compared using the area under ROC curves.	False negative=8* True negative=99* Sensitivity=56 (95% CI: 34-75)* Specificity=91 (95% CI: 87-94)* Positive LR= 6.06 (95% CI: 2.65-12.32)* Negative LR= 0.49 (95% CI: 0.27-0.76)* AUROC NR APRI versus published criteria for CFLD diagnosis n=122 Sample size reported do not match with the reported number of patients with and without CFLD. >0.4 True positive=9* False positive=9* False negative=9* True negative=100* Sensitivity=50 (95% CI: 29-69)* Specificity=92 (95% CI: 88-95)* Positive LR= 6.06 (95% CI: 2.48-13.50)*	Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Unclear 3.A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3.B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW RISK Flow and Timing Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? No Were all patients included in the analysis? No 4.A Could the patient flow have introduced bias? HIGH RISK Other information

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Negative LR= 0.55 (95% CI: 0.33-0.80)* AUROC=0.70 (95% CI: 0.54-0.86) >0.5** True positive=9* False positive=7* False negative=9* True negative=102* Sensitivity=50 (95% CI: 29-68)* Specificity=94 (95% CI: 90-97)* Positive LR= 7.79 (95% CI: 2.99-19.44)* Negative LR= 0.53 (95% CI: 0.33-0.78)* AUROC NR FibroTest versus published criteria for CFLD diagnosis n=106 >0.10** True positive=14* False positive=38* False negative=3* True negative=51* Sensitivity=82 (95% CI: 58-95)* Specificity=57 (95% CI: 53-60)*	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Positive LR= 1.93 (95% CI: 1.22-2.37)* Negative LR= 0.31 (95% CI: 0.08-0.80)* AUROC=0.76 (95% CI: 0.62-0.90) >0.20 True positive=6* False positive=10* False negative=11* True negative=79* Sensitivity=35 (95% CI: 16-53)* Specificity=89 (95% CI: 85-93)* Positive LR= 3.14 (95% CI: 1.10-7.80)* Negative LR= 0.73 (95% CI: 0.47-0.98)* AUROC NR *Calculated by the NGA technical team from data reported in the article **Optimal cut-offs of tests defined by the maximal sum of sensitivity and specificity LR = likelihood ratio NR= not reported	
Full citation	Sample size n=66	Tests Reference standard	Methods	Results	Limitations QUADAS 2 checklist

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Witters, P., De Boeck, K., Dupont, L., Proesmans, M., Vermeulen, F., Servaes, R., Verslype, C., Laleman, W., Nevens, F., Hoffman, I., Cassiman, D., Non-invasive liver elastography (Fibroscan) for detection of cystic fibrosis-associated liver disease, Journal of Cystic Fibrosis, 8, 392-9, 2009 Ref Id 330202 Country/ies where the study was carried out Belgium Study type Prospective cohort study Aim of the study To evaluate the diagnostic accuracy compared to other diagnostic tools as well as the relation of the liver stiffness	Characteristics The CF group (n = 66) consisted of 36 male and 30 female patients with a mean age of 13.6 ± 7.8 yr (32 patients < 12 yr, 24 patients between 12 and 18 yr and 10 patients > 18 yr. Six patients (9%) had evidence of clinical CFLD (hepatomegaly or splenomegaly) and 7 (11%) had evidence of biochemical CFLD. Ultrasonography revealed hepatomegaly in 15 (23%) patients and splenomegaly in 16 patients (24%). 26 patients (39%) had clinical, biochemical or ultrasonographic CFLD. A control group with no liver disease (n = 59) consisted of 26	The North-American cystic fibrosis foundation (CFF) consensus workgroup defines CFLD as the presence of either clinical or biochemical liver disease. -Clinical liver disease was defined as the presence of hepatomegaly or splenomegaly -Biochemical liver disease was defined as the elevation of 2 of these tests: Liver tests (AST, ALT, alkaline phosphatase, bilirubin and gamma-GT) from all CF patients from January 1996 to July 2007 were studied and patients with persistently elevated liver tests were identified (3–6 months, 1.5 times age-dependent upper limit of normal). Index tests -FibroScan: Liver stiffness was assessed by transient elastography (Fibroscan, Echosens, Paris). At least 10	Setting: CF clinic at the university hospital Fibroscan measurements were performed in 66 CF patients. Age-specific cutoff values were determined in a control population (n = 59) and was set at 5.63kPa for <12 years and 6.50kPa for ≥12 years. The measurements were compared to clinical data, biyearly biochemistry and ultrasound.	Ultrasound versus clinical CFLD definition in detection of CFLD n=66 patients True positive=4* False positive=20* False negative=2* True negative=40* Sensitivity=66.7 (95% CI: 25.0-93.9)* Specificity=66.7 (95% CI: 62.5-69.4)* Positive LR= 2.0 (95% CI: 0.67-3.07)* Negative LR= 0.50 (95% CI: 0.09-1.2)* AUROC=0.77 (95% CI: 0.51-1.02) Ultrasound versus biochemical CFLD definition in detection of CFLD n=66 patients True positive=3* False positive=20* False negative=3* True negative=40* Sensitivity=50.0 (95% CI: 14.3-85.6)* Specificity=66.7 (95% CI: 63.1-70.2)* Positive LR= 1.5 (95% CI: 0.39-2.88)*	Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Unclear 1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes for Clinical,biochemical and ultrasound index tests and was determined for Fibroscan using a control population 2.A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2.B Is there concern that the index test, its conduct, or interpretation differ from the

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
to risk factors for CFLD Study dates 1996-2007 Source of funding Not reported	male and 33 female subjects with a mean age of 10.2 ± 3.7 yr (41 patients < 12 yr, 18 patients 12–18 yr) and were investigated to define normal values of liver stiffness only. Inclusion Criteria -CF patients followed up in CF clinic at a university hospital: Clinical liver disease was defined as the presence of hepatomegaly or splenomegaly Exclusion Criteria Not reported	measurements per patient are obtained, using the standard probe. Median values and interquartile range (IQR, kPa) are reported. A success-rate of at least 60% was considered necessary. In the paediatric population special care was taken in order to make sure there was no A-shaped wave on the elastogram which indicates an incorrectly accepted (non-automatically rejected) measurement leading to an overestimation of the stiffness produced by influence of the surrounding rib bone and soft tissue. Fibroscan liver disease was defined as a result above the age-related upper limit of normal liver stiffness. - Ultrasound: Ultrasonographic liver disease was defined as a Williams score of at least 4/9 (i.e. intermediate coarse to		Negative LR= 0.75 (95% CI: 0.21-1.36)* AUROC=0.62 (95% CI: 0.40-0.84) Ultrasound versus CFF consensus definition in detection of CFLD n=66 patients True positive=7* False positive=16* False negative=4* True negative=39* Sensitivity=63.6 (95% CI: 33.6-87.0)* Specificity=70.9 (95% CI: 64.9-75.6)* Positive LR= 2.19 (95% CI: 0.96-3.56)* Negative LR= 0.51 (95% CI: 0.17-1.02)* AUROC=0.70 (95% CI: 0.51-0.89) Fibroscan versus clinical CFLD definition in detection of CFLD n=66 patients True positive=5* False positive=9* False negative=1* True negative=51*	review question? LOW CONCERN Reference Standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Unclear 3.A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3.B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN Flow and Timing Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4.A Could the patient flow have introduced bias? LOW RISK

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		irregular liver parenchyma, liver edge nodularity and/or moderate to severe periportal fibrosis).		Sensitivity=83.3 (95% CI: 38.7-99.1)* Specificity=85.0 (95% CI: 80.5-86.6)* Positive LR= 5.6 (95% CI: 2.0-7.4)* Negative LR= 0.20 (95% CI: 0.01-0.76)* AUROC=0.93 (95% CI: 0.85-1.01) Fibroscan versus biochemical CFLD definition in detection of CFLD n=66 patients True positive=3* False positive=10* False negative=3* True negative=50* Sensitivity=50.0 (95% CI: 14.5-85.3)* Specificity=83.3 (95% CI: 79.8-86.9)* Positive LR= 3.0 (95% CI: 0.72-6.5)* Negative LR= 0.60 (95% CI: 0.17-1.07)* AUROC=0.78 (95% CI: 0.61-0.95) Fibroscan versus CFF consensus definition in detection of CFLD	Other information

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				n=66 patients True positive=7* False positive=7* False negative=4* True negative=48* Sensitivity=63.6 (95% CI: 34.4-86.0)* Specificity=87.3 (95% CI: 81.4-91.8)* Positive LR= 5.0 (95% CI: 1.86-10.43)* Negative LR= 0.42 (95% CI: 0.15-0.81)* AUROC=0.86 (95% CI: 0.74-0.98) *Calculated by the NGA technical team from data reported in the article LR = likelihood ratio	
Full citation Lemaitre, C., Dominique, S., Billoud, E., Eliezer, M., Montialoux, H., Quillard, M., Riachi, G., Koning, E., Morisse-Pradier, H., Savoye, G., Savoye-Collet, C., Goria, O., Relevance of 3D Cholangiography and Transient	Sample size N=25 (out of cohort of 64) Characteristics of studied patients were not statistically different compared to the whole CF population. Characteristics	Tests Index test: Transient Elastography LSM by transient elastography was measured by Fibroscan(Echosens, Paris, France, size M) Ten measurements were taken in 3 different sites, and results are expressed	Methods Design: retrospective one-year cross-sectional cohort study Setting: cystic fibrosis reference centre at Rouen University Hospital Procedure: clinical and genetic characteristics were	Results Transient elastography versus liver function test or ultrasound for detection of CFLD n=23 True positive=3* False positive=3* False negative=1* True negative=16* Sensitivity=75 (95% CI: 24.2-98.6)*	Limitations QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? No Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Unclear

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Elastography to Assess Cystic Fibrosis-Associated Liver Disease?, Canadian Respiratory JournalCan Respir J, 2016, 4592702, 2016 Ref Id 537183 Country/ies where the study was carried out France Study type Retrospective cohort study Aim of the study To describe the usefulness of magnetic resonance imaging (MRI) and liver stiffness measurement (LSM) for the assessment of CFLD. Study dates Between July 2009 and July 2010. Source of funding Not reported.	Median age, years (range): 25 (18 to 43) Gender (M/F): 0.46 MI: 16% (n=4) Pancreatic insufficiency: 88% (n=22) UDCA treatment: 40% (n+10) FEV1%> 67.6% (50.4 to 84.8) Inclusion Criteria All adult patients with CF, investigated by hepatobiliary MRI and by transient elastography for liver stiffness measurement (LSM) between July 2009 and July 2010 Exclusion Criteria Patients in whom CFTR-related disorder was limited to one-organ dysfunction (i.e., congenital bilateral absence of vas deferens).	as a mean of 10 valid measurements Results were expressed in kilopascal (kPa) using the Metavir scoring system based on previous study of transient elastography in chronic biliary disease (Corpechot 2006): Metavir F0-F1 score corresponded to LSM of ¡Ý7.2 kPa, and F2, F3, and F4 corresponded to ¡Ý7.3 kPa, 9.8 kPa, and 17.3 kPa, respectively Index test: Biliary and Hepatic Magnetic Resonance Imaging Performed with 1.5 Tesla (Philips Achieva, Philips Medical Systems, Best, Netherlands) The following sequences were performed: (1) T1-weighted sequence, axial image (TR 183ms, TE 2.3ms, FOV 70 mm, slice thickness 7 mm, angle 55¡ã, 152 ¡Á 432); (2) T2-weighted	retrospectively collected from patient charts biochemical analysis (LFT, platelet counts, prothrombin time, albumin, and renal function) and routine abdominal US results including hepatic dysmorphia or PHT signs were also collected In all patients with abnormal LFT (any test > twice the normal values), additional workup was available including search for hepatitis B, hepatitis C, ferritin, transferrin saturation, and fasting lipid profile. Pulmonary function was collected, including forced expiratory volume. Statistical analysis Statistical analysis was conducted using SAS software version 9.3 SI units were used for all laboratory values with data summarized using mean ± standard deviation (SD) for continuous variables	Specificity=84.2 (95% CI: 73.5-69.2)* Positive LR= 4.75 (95% CI: 0.91-9.12)* Negative LR= 0.30 (95% CI: 0.02-1.03)* AUROC=NR MRI versus liver function test or ultrasound for detection of CFLD n=23 True positive=4* False positive=2* False negative=7* True negative=10* Sensitivity=36.4 (95% CI: 14.7-51.1)* Specificity=83.3 (95% CI: 63.5-96.8)* Positive LR= 2.18 (95% CI: 0.40-16.06)* Negative LR= 0.76 (95% CI: 0.50-1.34)* AUROC=NR *Calculated by the NGA technical team from data reported in the article NR=not reported LR = likelihood ratio	1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? Yes (the radiologist were blinded) If a threshold was used, was it pre-specified? Yes 2.A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN Reference Standard Is the reference standard likely to correctly classify the target condition? Unclear Were the reference standard results interpreted without knowledge of the results of the index test? Unclear 3.A Could the reference standard, its conduct, or its

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		sequence, axial SPAIR (TR 4459ms, TE 70ms, FOV 76 mm, slice thickness 6 mm, angle 90¡ã, 218 ¡Á 320); (3) T2-weighted sequence, axial HR (TR 1573ms, TE 100ms, FOV 79 mm, slice thickness 7 mm, angle 90¡ã, 341 ¡Á 560); (4) T2-weighted sequence diffusion 2b (TR 1489ms, TE 59ms, FOV 90 mm, slice thickness 6 mm, 92 ¡Á 67); (5) 3D MR cholangiogram (TR 1341ms, TE 574ms, FOV 100 mm, slice thickness 2.4 mm, angle 90¡ã, 221 ¡Á 560); (6) in and out phase sequence (TR 175ms, TE 2.3ms (in), 4.8ms (out), FOV 40¡ã, slice thickness 4 mm, angle 80¡ã, 224 ¡Á 192). Radiologists (CSC and EK) reviewed all MRI results blinded to clinical or biochemical parameters and reached decisions by consensus.	Number (%) for all recorded categorical variables describing the study population LSM are expressed in kPa as median (IQR) Student's t-test was used to compare continuous variables Chi square test was used when comparing categorical variables To assess the diagnostic performance of LSM for prediction of PHT, the area under the receiver operating curve (AUROC) was calculated. Optimal LSM for prediction of PHTwas identified by estimating sensitivity and specificity for various cut-offs. Prevalence of abnormalities in MRI and LSM was compared regarding the presence or not of LFT and/or US abnormalities using chi-square test and Fisher's exact test.		interpretation have introduced bias? UNCLEAR 3.B Is there concern that the target condition as defined by the reference standard does not match the review question? UNCLEAR Flow and Timing Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? 2 participants missing 4.A Could the patient flow have introduced bias? UNCLEAR Other information Conflict of interest: none.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
		The following items were studied for each patient using a standardized scale: atrophy of either right or left hepatic lobe and/or hypertrophy of the caudate lobe, marked lobulations of liver surface, first-segment hypertrophy, splenomegaly (long axis superior to 12 cm), portal vein dilatation (diameter superior to 12 mm), splenic vein dilatation, intrahepatic or extrahepatic biliary duct irregularity (segmental strictures and dilatations), ascites, and steatosis. Reference standard: liver function test or ultrasound Details not reported			
Full citation Woodruff, S. A., Sontag, M. K., Accurso, F. J., Sokol, R. J., Narkewicz, M. R., Prevalence of elevated liver enzymes in children	Sample size N=298 children with CF identified by newborn screening. Characteristics Method of diagnosis	Tests Monitoring strategy based on the assessment of liver function tests.	Methods Procedure: Clinical and laboratory data were collected prospectively. AST, ALT and GGT was obtained. Children were seen twice per year for the	Results Prognostic value of AST - Hazards ratio (95% CI): ≥1.5× ULN: 6.53 (2.02–21.1) ≥2.0× ULN: 6.52 (0.72–138.5)	Limitations The quality of this study was assessed using the tool proposed by Hayden et al. (2006), as suggested by NICE methods manual (2014) (full citation: Hayden JA, Cote P, Bombardier C (2006)

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
with cystic fibrosis diagnosed by newborn screen, Journal of Cystic Fibrosis, no pagination, 2016 Ref Id 566881 Country/ies where the study was carried out USA Study type Prospective cohort study Aim of the study To determine the prognostic value to elevated liver enzymes in children with CF diagnoses by newborn screening. Study dates 1982 to 2005 Source of funding Not reported.	Newborn screening 240 (80.5%) Meconium ileus 42 (14.1%) Missed on newborn screen 16 (5.4%) Age at diagnosis, weeks (median, IQR) 3.8 (2.4–5.7) Male gender (N, %) 147 (49.3) Hispanic ethnicity 35 (11.7%) CFTR mutation severity Severe (2 classes 1–3) 209 (76.3%) Milder (at least 1 class 4 or 5) 24 (8.8%) Unknown 15 (14.9%) Inclusion Criteria All children with CF born in Colorado from 1982 to 2005, diagnosed by newborn screening, the presence of meconium ileus, or who were missed by		first 2 years of life and then annually. UDA data was not available in the database before 2005, so authors developed a standardized evaluation and management pathway, that included starting ursodeoxycholic acid therapy at 10–20 mg/kg/day only if AST, ALT or GGT were ≥2× the upper limit of normal for age (ULN) for ≥6 months or if there was clinical evidence of advanced liver disease (e.g., splenomegaly, firm hepatomegaly or complications of portal hypertension) from 1990 forward. Pancreatic enzyme replacement therapy was initiated on all infants at diagnosis and continued unless there was verification of pancreatic sufficiency. The authors followed CF Foundation guidelines for nutritional and pulmonary therapies. ALT was determined at annual well CF visits starting in 1982 with	Prognostic value of ALT Hazards ratio (95% CI): ≥1.5× ULN: 1.95 (0.81–4.27) ≥2.0× ULN: 1.88 (0.82–3.91) Prognostic value of GGTP - Hazards ratio (95% CI): ≥1.5× ULN: 4.03 (1.15–13.45) ≥2.0× ULN: 2.44 (0.86-6.13) Hazards Ratios for the presence of clinically diagnosed liver disease, adjusted for sex, CFTR mutation severity, and the presence of meconium ileus.	Evaluation of the quality of prognosis studies in systematic reviews. Annals of Internal Medicine 144: 427–37) 1. The study sample represents the population of interest with regard to key characteristics, sufficient to limit potential bias to the results. YES 2. Loss to follow-up is unrelated to key characteristics (that is, the study data adequately represent the sample), sufficient to limit potential bias. YES 3. The prognostic factor of interest is adequately measured in study participants, sufficient to limit potential bias. YES 4. The outcome of interest is adequately measured in study participants, sufficient to limit potential bias. YES 5. Important potential confounders are appropriately accounted for, limiting potential bias with respect to the prognostic factor of interest. YES 6. The statistical analysis is appropriate for the design of the study, limiting potential for

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
	newborn screening. A sweat chloride N60 mmol/L or two pathologic CFTR mutations consistent with CF were considered positive evidence of CF. Exclusion Criteria Not reported.		subsequent inclusion of AST and GGT in 1990. Values were classified as normal, elevated (any elevation above the ULN), ≥1.5× ULN, ≥2× ULN and ≥3× ULN based on normal values for age and sex at the time of their determination. Statistical analysis: Product-Limit Survival Estimates were used to assess the age at first abnormality for AST, ALT and GGT. Early liver enzyme elevation (defined as present before 5 years of age) and persistent elevation defined as 2 or more abnormal values obtained at least 6 months apart at the annual visits. Univariate relative risks were calculated for persistent elevation (for ≥1.5× and 2×ULN) with the presence of meconium ileus, sex, CFTR mutation severity and Hispanic ethnicity. Due to missing values in children who did not have an 'annual visit'		the presentation of invalid results. YES OVERALL QUALITY: HIGH Other information Conflict of interest: 1 author was a consultant at Vertex. No other interest to declare.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
			recorded, data from all visits were used, and the mean clinical outcome for each year of age (rounded to the nearest year of age) was calculated. Clinical advanced liver disease was defined as the presence of cirrhosis (by imaging or liver histology), portal hypertension (by the presence of ascites, splenomegaly or thrombocytopenia, esophageal or gastric varices, or portal hypertensive gastropathy) or stage 3/4 fibrosis on liver biopsy obtained for clinical indications.Statistical significance was assessed by using an α = 0.05. SAS 9.2 (Carey, NC) was used for all analyses.		