Assessment of hepatic steatosis by controlled attenuation parameter using the M and XL probes: an individual patient data meta-analysis

David Petroff, Valentin Blank*, Philip N Newsome, Shalimar, Cosmin Sebastian Voican, Maja Thiele, Victor de Lédinghen, Stephan Baumeler, Wah Kheong Chan, Gabriel Perlemuter, Ana-Carolina Cardoso, Sandeep Aggarwal, Magali Sasso, Peter J Eddowes, Michael Allison, Emmanuel Tsochatzis, Quentin M Anstee, David Sheridan, Jeremy F Cobbold, Sylvie Naveau, Monica Lupsor-Platon, Sebastian Mueller, Aleksander Krag, Marie Irles-Depe, David Semela, Grace Lai-Hung Wong, Vincent Wai-Sun Wong, Cristiane A Villela-Nogueira, Harshit Garg, Olivier Chazouillères, Johannes Wiegand, Thomas Karlas*

Summary

Background Diagnostic tools for liver disease can now include estimation of the grade of hepatic steatosis (S0 to S3). Controlled attenuation parameter (CAP) is a non-invasive method for assessing hepatic steatosis that has become available for patients who are obese (FibroScan XL probe), but a consensus has not yet been reached regarding cutoffs and its diagnostic performance. We aimed to assess diagnostic properties and identify relevant covariates with use of an individual patient data meta-analysis.

Methods We did an individual patient data meta-analysis, in which we searched PubMed and Web of Science for studies published from database inception until April 30, 2019. Studies reporting original biopsy-controlled data of CAP for non-invasive grading of steatosis were eligible. Probe recommendation was based on automated selection, manual assessment of skin-to-liver-capsule distance, and a body-mass index (BMI) criterion. Receiver operating characteristic methods and mixed models were used to assess diagnostic properties and covariates. Patients with non-alcoholic fatty liver disease (NAFLD) were analysed separately because they are the predominant patient group when using the XL probe. This study is registered with PROSPERO, CRD42018099284.

Findings 16 studies reported histology-controlled CAP including the XL probe, and individual data from 13 papers and 2346 patients were included. Patients with a mean age of 46·5 years (SD 14·5) were recruited from 20 centres in nine countries. 2283 patients had data for BMI; 673 (29%) were normal weight (BMI <25 kg/m²), 530 (23%) were overweight (BMI ≥25 to <30 kg/m²), and 1080 (47%) were obese (BMI ≥30 kg/m²). 1277 (54%) patients had NAFLD, 474 (20%) had viral hepatitis, 285 (12%) had alcohol-associated liver disease, and 310 (13%) had other liver disease aetiologies. The XL probe was recommended in 1050 patients, 930 (89%) of whom had NAFLD; among the patients with NAFLD, the areas under the curve were 0·819 (95% CI 0·769–0·869) for S0 versus S1 to S3 and 0·754 (0·720–0·787) for S0 to S1 versus S2 to S3. CAP values were independently affected by aetiology, diabetes, BMI, aspartate aminotransferase, and sex. Optimal cutoffs differed substantially across aetiologies. Risk of bias according to QUADAS-2 was low.

Interpretation CAP cutoffs varied according to cause, and can effectively recognise significant steatosis in patients with viral hepatitis. CAP cannot grade steatosis in patients with NAFLD adequately, but its value in a NAFLD screening setting needs to be studied, ideally with methods beyond the traditional histological reference standard.

Funding The German Federal Ministry of Education and Research and Echosens.

Introduction

Fatty liver, the hepatic manifestation of the metabolic syndrome, has become an important field of clinical research, and the development of fast and easily applicable non-invasive methods for assessment of fatty liver disease is of great value.[1] Serum-based markers and liver stiffness measurement are now established and valuable tools to assess the degree of hepatic fibrosis.[1,2] In the past 5 years, non-invasive quantification of hepatic steatosis has attracted scientific interest and the continued development of technologies will help define the clinical implications of steatosis.[3–7] MRI methods permit precise estimation of total hepatic fat, which is of value in clinical studies,[8] but is not ideal for the screening of large populations at risk.[9] By contrast, the ultrasound-based controlled attenuation parameter (CAP) technology allows estimation of hepatic fat content during liver stiffness measurement with vibration-controlled transient elastography on FibroScan equipment (Echosens, Paris, France).[10] CAP was originally available only with the M probe (appropriate for lean individuals who typically have a skin-to-liver-capsule distance [SLD] of <25 mm)

Lancet Gastroenterol Hepatol 2021

Published Online January 15, 2020 https://doi.org/10.1016/S2468-1253(20)30357-5

See Online/Comment https://doi.org/10.1016/S2468-1253(20)30367-8

*Contributed equally

Clinical Trial Centre (D Petroff PhD) and Faculty of Medicine, Integrated Research and Treatment Center AdiposityDiseases, (D Petroff, V Blank MD), University of Leipzig, Leipzig, Germany; Division of Gastroenterology (V Blank, T Karlas MD) and Division of Hepatology (J Wiegand MD), Department of Medicine II, Leipzig University Medical Center, Leipzig, Germany; National Institute for Health Research, Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust (Prof P N Newsome PhD, P J Eddowes MD) and Centre for Liver Research, Institute of Immunology and Immunotherapy (Prof P N Newsome, P J Eddowes), University of Birmingham, Birmingham, UK; Liver Unit, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK (Prof P N Newsome, P J Eddowes); Department of Gastroenterology (Shalimar MD) and Department of Surgical Disciplines (H Garg MD, Prof S Aggarwal MD), All India Institute of Medical Sciences, New Delhi, India; Faculté de Médecine Paris-Sud, Université Paris-Sud, Université Paris-Saclay, Le Kremlin-Bicêtre,

France (C S Voican PhD, Prof G Perlemuter MD, Prof S Naveau MD)**; Service d'Hépato-Gastroentérologie et Nutrition, Hôpital Antoine-Béclère, Hôpitaux Universitaires Paris-Sud, Assistance Publique-Hôpitaux de Paris, Clamart, France** (C S Voican, Prof G Perlemuter, Prof S Naveau)**; INSERM U996, DHU Hepatinov, Labex LERMIT, Clamart, France** (C S Voican, Prof G Perlemuter, Prof S Naveau)**; Department of Gastroenterology and Hepatology, Odense University Hospital, University of Southern Denmark, Odense, Denmark** (M Thiele PhD, Prof A Krag PhD)**; Centre d'Investigation de la Fibrose Hépatique, Bordeaux University Hospital, Pessac, France** (Prof V de Lédinghen MD, M Irles-Depe MD)**; INSERM U1053, Bordeaux University, Bordeaux, France** (Prof V de Lédinghen, M Irles-Depe)**; Department of Gastroenterology and Hepatology, Cantonal Hospital St Gallen, St Gallen, Switzerland** (S Baumeler MD, D Semela MD)**; Gastroenterology and Hepatology Unit, Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia** (Prof W K Chan PhD)**; Hepatology Unit, Hospital Universitário Clementino Fraga Filho, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil** (A-C Cardoso MD, Prof C A Villela-Nogueira MD)**; Research and Development Department, Echosens, Paris, France** (M Sasso PhD)**; National Institute for Health Research Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust, Nottingham, UK** (P J Eddowes)**; University of Nottingham, Nottingham, UK** (P J Eddowes)**; Liver Unit, Addenbrooke's Hospital, Cambridge Biomedical Research Centre, Cambridge, UK** (M Allison MD)**; University College London Institute for Liver and Digestive Health, Royal Free Hospital, London, UK** (Prof E Tsochatzis PhD)**; Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK** (Prof Q M Anstee MD);

Research in context

Evidence before this study

Original research and meta-analyses showed that non-invasive grading of hepatic steatosis with ultrasound-based controlled attenuation parameter (CAP) is accurate for various liver diseases when using the M probe. A limitation of this probe is that it is not appropriate for patients who are obese (body-mass index [BMI] ≥30 kg/m²) meaning that its application in non-alcoholic fatty liver disease (NAFLD) was limited. In the past decade, the XL probe was introduced, which was designed for patients with a higher BMI. Several studies were done using CAP with the XL probe and showed promising results for steatosis grading compared with the histological gold-standard. By contrast, a recent conventional meta-analysis investigated diagnostic properties in NAFLD and concluded that diagnostic capabilities should be viewed with caution. However, only two of the nine studies included in that meta-analysis used the XL probe. Consensus regarding optimal cutoffs was not available and uncertainties in diagnostic performance and relevant covariates led us to do an individual patient data meta-analysis. In a systematic literature search, we identified 16 studies with histology controlled CAP that included the XL probe. We were able to obtain individual patient data from 13 of them and 2346 patients were included in the final analysis.

Added value of this study

The diagnostic performance and optimal cutoffs for CAP with the XL probe depend substantially on aetiology, diabetes status, BMI, and sex. The area under the curve (AUC) from receiver operating characteristic analysis was only 0·754 for distinguishing S2 to S3 from S0 to S1, and 0·717 for distinguishing S3 from S0 to S2 in patients with NAFLD. The AUC for detection of any steatosis (S1 to S3) is better at 0·819, but S0 cases are heavily influenced by patients scheduled for bariatric surgery. Overall, the diagnostic performance of CAP in patients undergoing bariatric surgery is better than for the typical patient with NAFLD. Previously proposed reliability criteria do not lead to improved performance. Comparison of M and XL probe measurements in patients for whom both probes are viable (BMI >23 and <30 kg/m²) shows high variance for the difference between measurements, although the mean difference is small.

Implications of all the available evidence

The interpretation of CAP should take liver disease cause and covariates into account. Even with the XL probe, CAP cannot grade steatosis in patients with NAFLD adequately. The findings in this study should be considered when developing or interpreting non-invasive diagnostics for NAFLD and non-alcoholic steatohepatitis with methods that include CAP. Its value in a NAFLD screening setting needs to be studied, ideally with methods beyond the traditional histological reference standard such as MRI.

and discriminates fairly well between histological grades of steatosis in patients with viral hepatitis and non-alcoholic fatty liver disease (NAFLD), as shown in an individual patient data meta-analysis by our group.[11] The CAP algorithm later introduced for the XL probe facilitates steatosis assessment in patients who are obese (SLD ≥25 mm),[12] in whom steatosis assessment is most needed. Several biopsy controlled studies have assessed correlation of XL-CAP with histological steatosis grades, especially in patients who are obese (BMI ≥30 kg/m²) with NAFLD.[12–27]

The availability of both probes could provide practitioners with the necessary tools to grade steatosis quickly and non-invasively in patients at risk, as shown in one multicentre, prospective study on patients undergoing liver biopsy for suspicion of NAFLD or non-alcoholic steatohepatitis (NASH).[14] However, an in-depth meta-analysis considering all relevant covariates and reliability criteria is necessary to account for cohorts with different causes of liver disease and broader inclusion criteria. Moreover, larger cohorts are needed to evaluate diagnostic performance and ascertain whether or not cutoffs established with the M probe are appropriate for the XL probe, especially considering that the M probe might have been used inappropriately in patients who are overweight (BMI ≥25 kg/m²).[28] To assess its performance and provide a comprehensive comparison with the M probe, we aimed to do an extensive individual patient data meta-analysis of studies that included use of the XL probe.

Methods

Search strategy and selection criteria

We did an individual patient data meta-analysis. Search terms and time intervals were defined at two consensus meetings of the study group in Paris, France on April 13, 2018, and Vienna, Austria on April 13, 2019. Only English language manuscripts published or accepted by a peer-reviewed journal up until April 30, 2019, qualified for inclusion in the meta-analysis. Study results published only in abstract form were not considered.

Publications on the diagnostic properties of CAP for liver steatosis quantification with histology as reference standard were extracted from PubMed and Web of Science from their inception until April 30, 2019. The systematic literature search used the term ''controlled attenuation parameter XL". The final query for PubMed was: "controlled"[All Fields] AND "attenuation"[All Fields] AND "parameter"[All Fields] AND "XL"[All Fields] AND "english"[Language] NOT "review"[Publication Type] AND 1900/1/1:2019/4/30[Date - Entry]. An analogous query was used for Web of Science. Titles and abstracts

were screened for eligibility by two authors (TK and VB). In case of disagreement, a third researcher (DP) was consulted to reach consensus. The articles selected for a full-text review were examined by two researchers (TK, DP) to establish whether or not the inclusion and exclusion criteria were met.

In addition to this search strategy, we screened abstract books of the International Liver Congress (European Association for the Study of the Liver, 2016–19) and The Liver Meeting (American Association for the Study of Liver Diseases, 2016–18) using the same search terms as above. The corresponding authors were contacted and invited to participate in the meta-analysis if their full papers were accepted within the pre-defined time limit.

Studies reporting original biopsy-controlled data of CAP for non-invasive grading of steatosis were eligible for inclusion in the individual patient data meta-analysis. CAP data for the vibration-controlled transient elastography XL probe (Echosens, Paris, France), which became commercially available in 2014, had to be provided in the paper. Histology specimens must have been evaluated for steatosis according to the percentage of hepatocytes involved.[29] Approval of the local ethics committee was required and had to be forwarded on to the analysis team. Studies that selected patients based primarily on the result of a CAP measurement were not included in the individual patient data meta-analysis to avoid selection bias. Studies were also excluded if data could not be obtained despite multiple attempts to contact the study investigators. If individual data sets were published repeatedly, only the latest data set was used. This study was approved by the ethics committee of the coordinating site in Leipzig (218/15-ek).

Individual data verification and study quality assessment

The quality of each study was appraised by two authors (VB and TK for half the articles and VB and DP for the other half) with use of the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2).[30] Discrepancies were discussed and resolved by these three authors. As an important further test of data quality and to ensure that they were correctly interpreted, individual data from each study were used to reproduce the main results. Inconsistencies were communicated to the corresponding authors of the studies and resolved via discussions and revised data sets, as necessary.

Steatosis grading and fibrosis staging

Histopathological data were used as reference standard for the evaluation of CAP. Steatosis was defined according to the number of affected hepatocytes: S0 (<5%), S1 (5–33%), S2 (34–66%), and S3 (>66%).[29] Fibrosis was staged according to the METAVIR score (for viral hepatitis) or NAFLD Activity Score (for NAFLD and alcohol-associated liver disease).[29,31] Quality criteria for the histological specimens (ie, number of portal tracts and length of specimen) differed between studies; however, the risk of sampling variability is low for steatosis grading.[32] We therefore excluded only samples that were classified as unreliable by the respective study pathologist. Because steatosis grade can be modulated by lifestyle and medical interventions, individual patients were excluded if the time interval between biopsy and CAP measurement was more than 6 months, as stipulated in the protocol and ratified at the study group meetings.

Vibration-controlled transient elastography including CAP measurement

CAP acquisition took place in all studies during liver stiffness measurement with the vibration-controlled transient elastography device FibroScan in line with guideline recommendations.[2,33] Cases with at least ten single measurements for calculation of the CAP median were required. The manufacturer recommended using SLD at the measuring site for choice of the appropriate probe and incorporated an automated probe selection software in the latest model of the device. However, different methods of probe selection have been proposed[2,33] and were expected among the studies. Therefore, we used an approach to define the so-called correct probe choice on the basis of availability of data. These were automated probe selection, manual assessment of SLD, but also a body-mass index (BMI) criterion proposed in international guidelines (cutoff 30 kg/m²).[2] If anthropometric data indicated a high risk of inappropriate probe application, the respective cases were excluded from the primary analysis. For the XL probe, individuals with a BMI of less than 18 kg/m² were not considered, and the M probe was not accepted in individuals with a BMI of more than 35 kg/m² unless the probe was selected by the internal software algorithm or SLD was less than 25 mm.[34] In all analyses of diagnostic performance, the data from the correct probe were always used. Following recommendations on the interpretation of liver stiffness measurement data for fibrosis estimation,[35] reliability criteria for CAP based on the IQR of the single CAP measurements have been proposed: in 2017 Wong and colleagues[36] suggested an IQR threshold of more than 40 dB/m for poorly reliable measurements with the M probe, whereas Semmler and colleagues[37] reported the CAP-IQR to CAP median ratios of less than 0·1, less than 0·2, and less than 0·3 as reliability indicators for both the M and XL probes. We therefore assessed the ability of these criteria to identify reliable CAP measurements.

Objectives

The primary objective was to establish optimal cutoffs distinguishing mild (≤S1) from advanced steatosis (S2 to S3) for the XL probe and healthy (S0) from affected (S1 to S3) patients, if the number of patients with S0 was sufficiently high. Secondary objectives were to estimate the probability of a given steatosis grade at any specific

Newcastle National Institute for Health Research Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK (Prof Q M Anstee)**; Institute of Translational and Stratified Medicine, Faculty of Medicine and Dentistry, University of Plymouth, Plymouth, UK** (D Sheridan MD)**; Department of Gastroenterology and Hepatology, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, UK** (J F Cobbold PhD)**; Department of Medical Imaging, Iuliu Hatieganu University of Medicine and Pharmacy, Regional Institute of Gastroenterology and Hepatology, Cluj-Napoca, Romania** (M Lupsor-Platon MD)**; Department of Medicine and Liver Diseases, Salem Medical Center, University of Heidelberg, Heidelberg, Germany** (Prof S Mueller MD)**; Institute of Digestive Disease** (Prof G L-H Wong MD, Prof V W-S Wong MD) **and Department of Medicine and Therapeutics** (Prof G L-H Wong, Prof V W-S Wong)**, The Chinese University of Hong Kong, Sha Tin, Hong Kong; Hepatology Department, Saint-Antoine Hospital, Assistance Publique–Hôpitaux de Paris, Centre de Recherche Saint-Antoine, Sorbonne University, Paris, France** (Prof O Chazouillères MD)

Correspondence to:
Dr Thomas Karlas, Division of Gastroenterology, Department of Medicine II, Leipzig University Medical Center, 04103 Leipzig, Germany
thomas.karlas@medizin.uni-leipzig.de

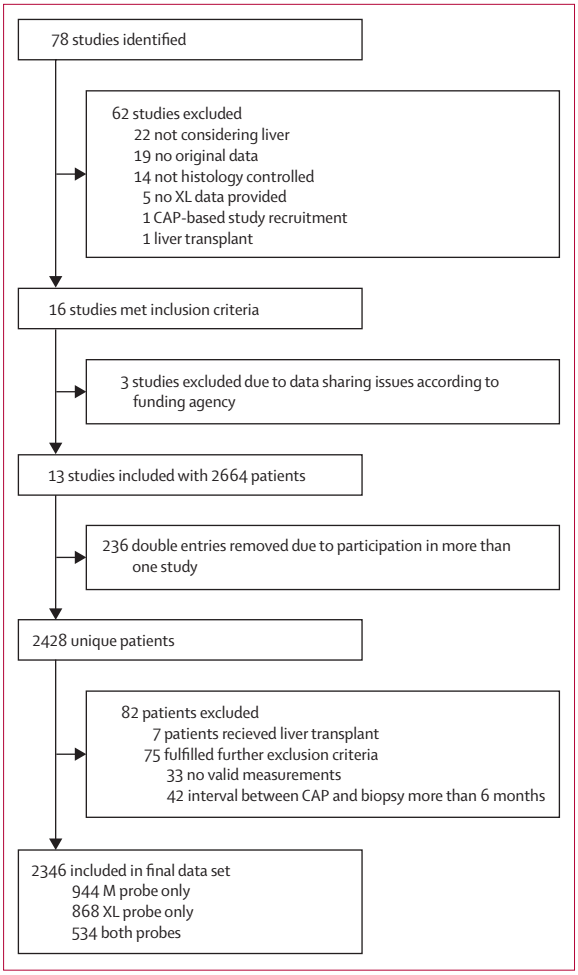

Figure 1: **Study selection**
CAP=controlled attenuation parameter.

See Online for appendix

CAP value, estimate the effects of covariates on CAP values, examine the use of established cutoffs, compare M and XL probes, and evaluate CAP quality criteria.

Data analysis

All analyses were done with the software R, version 3.6.1. Receiver operating characteristic (ROC) analyses including area under the curve (AUC) made use of the package OptimalCutpoints and the cutoff was found using the median Youden optimum from a bootstrap method with 10 000 samples.[38] For comparison with the literature, the rule-in and rule-out optimal cutoffs were chosen by requiring that their respective sensitivity (for rule-in) or specificity (for rule-out) be 0·9. CIs for variables based on proportions were constructed with Wilson score intervals[39] and for the optimal cutoffs, sensitivity and specificity from the quantiles of the bootstrapping samples. As stipulated in the protocol, a linear mixed model for CAP values included steatosis grade, BMI, diabetes status, sex, choice of probe, and cause of liver disease as covariates and the study as a random term and CIs were established with a profiling method.[40] These models were compared using ANOVA on the analogous fixed model (without study or with study as a fixed term). This permitted assessment of heterogeneity between studies from the variance of the random term, the coefficients in the study terms of the fixed effects model, and the difference in the remaining coefficients between models with and without study terms. Ordinal regression from the MASS package[40] was used to estimate the probability of each steatosis grade at a given CAP value after sampling the largest possible number of patients from the data with a prescribed prevalence of steatosis grades. Comparison of the M and XL probes was done with Bland-Altman methods. Two suggestions for CAP quality criteria were evaluated, one suggesting that the IQR of ten CAP measurements be smaller than 40 dB/m[36] and the other that the ratio of IQR to the median value be less than 0·1, 0·2, or 0·3.[37] These criteria were examined with the linear correlation coefficient between CAP and histologically determined percentage of affected hepatocytes at different quality thresholds.

This study was registered with PROSPERO, CRD42018099284.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, or preparation of the manuscript draft. Before submission of the manuscript, Echosens was presented a copy that could be commented on, but there was no obligation to abide by any suggestions made by the funder. The coauthors had final say on all aspects of the manuscript. This article was not commissioned by Echosens or any other company. All authors had full access to all the data and the final responsibility to submit for publication.

Results

Our original search yielded 78 papers, of which 16 met the study criteria (figure 1; appendix p 3). We were able to correspond with authors from all papers, but three did not provide data because of concerns on the part of the funding agency regarding data sharing. The remaining 13 papers comprised 2664 patients. 2346 (88%) of these patients met the inclusion criteria for the current analysis; patients were recruited from 20 centres in nine countries. Due to overlapping patients, two papers are listed in the following only as Chan et al, 2018.[22,26] Two papers provided more than 300 patients, six between 100 and 200 patients, and the remaining four provided less than 100 patients. CAP was done within 1 day of the reference test in 1717 (73%) patients and within 1 month in 2292 (97%) patients.

Patient characteristics are shown in table 1 and an assessment of the quality of the studies according to QUADAS-2[30] can be found in the appendix (p 2) showing that the risk of bias was low, but that patient selection and conduct of the index test could have introduced bias when applied to our research question.

	S0 (n=750)	S1 (n=623)	S2 (n=501)	S3 (n=472)	All patients (n=2346)
Sex					
Female	346 (46%)	306 (49%)	251 (50%)	244 (52%)	1147 (49%)
Male	404 (54%)	317 (51%)	250 (50%)	228 (48%)	1199 (51%)
Age, years	40·6 (15·4)	49·1 (14·0)	51·4 (12·5)	47·3 (12·6)	46·5 (14·5)
Body-mass index, kg/m²*	26·2 (7·7)	32·5 (8·9)	33·4 (8·2)	36·3 (8·4)	31·4 (9·1)
<25	425/732 (58%)	127/593 (21%)	75/489 (15%)	46/469 (10%)	673/2283 (29%)
≥25 to <30	168/732 (23%)	158/593 (27%)	127/489 (26%)	77/469 (16%)	530/2283 (23%)
≥30 to <35	41/732 (6%)	98/593 (17%)	102/489 (21%)	81/469 (17%)	322/2283 (14%)
≥35 to <40	39/732 (5%)	75/593 (13%)	77/489 (16%)	113/469 (24%)	304/2283 (13%)
≥40 to <45	35/732 (5%)	80/593 (13%)	59/489 (12%)	91/469 (19%)	265/2283 (12%)
≥45	24/732 (3%)	55/593 (9%)	49/489 (10%)	61/469 (13%)	189/2283 (8%)
Liver disease					
NAFLD or NASH	96 (13%)	376 (60%)	394 (79%)	411 (87%)	1277 (54%)
Hepatitis B	195 (26%)	44 (7%)	11 (2%)	8 (2%)	258 (11%)
Hepatitis C	130 (17%)	54 (9%)	21 (4%)	11 (2%)	216 (9%)
Alcohol-associated liver disease	76 (10%)	102 (16%)	69 (14%)	38 (8%)	285 (12%)
Other	253 (34%)	47 (8%)	6 (1%)	4 (1%)	310 (13%)
Type 2 diabetes†					
NAFLD or NASH	9/96 (9%)	156/361 (43%)	188/387 (49%)	160/409 (39%)	513/1253 (41%)
Hepatitis B	8/193 (4%)	2/43 (5%)	4/11 (36%)	2/8 (25%)	16/255 (6%)
Hepatitis C	9/127 (7%)	5/54 (9%)	3/21 (14%)	1/11 (9%)	18/213 (8%)
Alcohol-associated liver disease	6/76 (8%)	14/91 (15%)	11/68 (16%)	6/38 (16%)	37/273 (14%)
Other	41/222 (18%)	5/47 (11%)	2/5 (40%)	0/4 (0%)	48/278 (17%)
Alanine aminotransferase‡, U/L					
Female	33 (21–62)	30 (21–53)	35 (24–54)	47 (33–77)	36 (23–62)
Male	48 (29–89)	43 (28–67)	53 (35–83)	71 (47–110)	51 (33–86)
Aspartate aminotransferase§, U/L					
Female	30 (22–53)	30 (22–48)	30 (24–42)	42 (30–70)	33 (23–53)
Male	39 (28–57)	36 (26–55)	40 (28–61)	51 (37–70)	40 (29–60)
Time between CAP and biopsy					
CAP >14 days before biopsy	1 (<1%)	3 (<1%)	6 (1%)	10 (2%)	20 (1%)
CAP 2–14 days before biopsy	70 (9%)	116 (19%)	98 (20%)	173 (37%)	457 (19%)
CAP within 1 day of biopsy	667 (89%)	433 (70%)	345 (69%)	269 (57%)	1714 (73%)
CAP 2–14 days after biopsy	9 (1%)	49 (8%)	29 (6%)	9 (2%)	96 (4%)
CAP >14 days after biopsy	3 (<1%)	22 (4%)	23 (5%)	11 (2%)	59 (3%)
Liver stiffness, kPa					
M probe¶	6·6 (5·1–9·6)	8·2 (6·1–12·0)	8·8 (6·1–13·1)	9·4 (6·6–13·5)	7·8 (5·7–11·9)
XL probe‖	6·3 (4·9–8·8)	6·6 (5·0–9·9)	7·3 (5·4–10·7)	7·9 (5·8–12·4)	6·9 (5·3–10·4)
Fibrosis staging**					
F0	264/722 (37%)	130/604 (22%)	74/495 (15%)	60/461 (13%)	528/2282 (23%)
F1	208/722 (29%)	221/604 (37%)	182/495 (37%)	179/461 (39%)	790/2282 (35%)
F2	106/722 (15%)	113/604 (19%)	98/495 (20%)	106/461 (23%)	423/2282 (19%)
F3	66/722 (9%)	75/604 (12%)	89/495 (18%)	86/461 (19%)	316/2282 (14%)
F4	78/722 (11%)	65/604 (11%)	52/495 (11%)	30/461 (7%)	225/2282 (10%)

Data are n (%), mean (SD), or median (IQR). CAP=controlled attenuation parameter. NAFLD=non-alcoholic fatty liver disease. NASH=non-alcoholic steatohepatitis. Steatosis was defined according to the number of affected hepatocytes: S0 (<5%), S1 (5–33%), S2 (34–66%), and S3 (>66%). *Data available from 2283 patients. †Data available from 2272 patients. ‡Data available from 2307 patients. §Data available from 2282 patients. ¶Data available from 1478 patients. ||Data available from 1270 patients. **Data available from 2282 patients.

Table 1: **Population characteristics**

A high proportion of patients had NAFLD (1277 [54%]) and a high BMI (454 [20%] for BMI ≥40 kg/m²), many of whom were candidates for bariatric surgery. However, in the NAFLD cohort, only 185 (15%) of the patients had a BMI greater than or equal to 45 kg/m². 96 (8%) of 1277 patients with NAFLD were classified as S0, 72 of

	S0 *vs* S1 to S3	S0 to S1 *vs* S2 to S3	S0 to S2 *vs* S3
NAFLD XL probe*			
AUC	0·819 (0·769–0·869)	0·754 (0·720–0·787)	0·717 (0·684–0·751)
Prevalence	0·911	0·630	0·346
Optimal cutoff, dB/m	297 (287–323)	317 (306–334)	333 (320–340)
Sensitivity	0·798 (0·771–0·824)	0·783 (0·748–0·816)	0·761 (0·714–0·808)
Specificity	0·735 (0·639–0·831)	0·628 (0·574–0·679)	0·597 (0·559–0·636)
NAFLD†			
AUC	0·807 (0·758–0·855)	0·736 (0·707–0·765)	0·711 (0·682–0·740)
Prevalence	0·925	0·630	0·322
Optimal cutoff, dB/m	294 (286–313)	310 (305–321)	331 (319–340)
Sensitivity	0·790 (0·767–0·813)	0·790 (0·761–0·817)	0·718 (0·674–0·762)
Specificity	0·740 (0·646–0·823)	0·589 (0·557–0·634)	0·621 (0·589–0·652)
Hepatitis B or hepatitis C‡			
AUC	0·769 (0·724–0·814)	0·847 (0·794–0·900)	NA
Prevalence	0·311	0·106	0·040
Optimal cutoff, dB/m	230 (209–266)	264 (238–285)	NA
Sensitivity	0·714 (0·640–0·789)	0·760 (0·640–0·880)	NA
Specificity	0·680 (0·628–0·729)	0·794 (0·754–0·832)	NA
Alcohol-associated liver disease§			
AUC	0·765 (0·704–0·826)	0·766 (0·711–0·821)	0·802 (0·733–0·871)
Prevalence	0·732	0·377	0·134
Optimal cutoff, dB/m	274 (236–291)	268 (263–310)	291 (285–319)
Sensitivity	0·644 (0·577–0·707)	0·860 (0·794–0·925)	0·868 (0·763–0·974)
Specificity	0·790 (0·697–0·882)	0·554 (0·480–0·627)	0·679 (0·622–0·736)
Other¶			
AUC	0·687 (0·612–0·762)	NA	NA
Prevalence	0·184	0·032	0·013
Optimal cutoff, dB/m	244 (205–262)	NA	NA
Sensitivity	0·684 (0·561–0·807)	NA	NA
Specificity	0·659 (0·599–0·718)	NA	NA

Data are estimate (95% CI), unless specified. Only data from the correct probe choice were included. If the available data were too scant, then estimates were not provided (eg, only four patients with other liver disease aetiologies were S3). AUC=area under the curve. CAP=controlled attenuation parameter. NAFLD=non-alcoholic fatty liver disease. NA=not applicable as data are too scant. Steatosis was defined according to the number of affected hepatocytes: S0 (<5%), S1 (5–33%), S2 (34–66%), and S3 (>66%). *n=930, 95% BMI >30 kg/m². †n=1274, 73% XL, 76% BMI >30 kg/m². ‡n=472, 10% XL, 8% BMI >30 kg/m². §n=284: 8% XL, 20% BMI >30 kg/m². ¶n=309, 7% XL, 16% BMI >30 kg/m².

Table 2: **Estimated diagnostic properties of CAP for assessing histologically defined steatosis grade**

whom were being considered for bariatric surgery and 17 came from a screening setting. There was a high prevalence of significant fibrosis (F >1, 964 [42%]) which was unevenly distributed between those patients with (478 [38%] of 1262 patients) and without (486 [48%] of 1020) NAFLD. Among the patients with NAFLD, those with S0 did not have advanced fibrosis. The liver stiffness measurement values in table 1 based on the M and XL probes cannot be compared, as they derive largely from different groups of patients.

The results of ROC analyses are provided in table 2 by liver disease and for the XL probe alone in patients with NAFLD. Only data from the correct probe according to the definition stated were included. For the XL probe in patients with NAFLD, the optimal cutoff to distinguish S0 to S1 from S2 to S3 was 317 dB/m (95% CI 306–334) and only marginally higher than the value of 310 dB/m (305–321) found in the total cohort of patients with NAFLD, in which the XL probe was used for 930 (73%) patients. This optimal cutoff was substantially higher than for patients with viral hepatitis, or alcohol-associated liver disease (table 2). All but 15 of the patients with alcohol-associated liver disease came from a single study,[41] but the Youden optimisation chosen in table 2 complements other choices in that paper, which were optimised to have sensitivity or specificity of 0·9. The optimal cutoff in patients with alcohol-associated liver disease for S0 versus S1 to S3 was paradoxically higher than for S0 to S1 versus S2 to S3 (274 dB/m *vs* 268 dB/m), but the CIs were wide and the shifts in sensitivity and specificity suggested that their sum remained fairly constant over a large range of CAP values. The large differences in optimal cutoffs between causes of liver disease indicated that they should be considered separately in many analyses. For some analyses, we chose to focus on NAFLD, which is where the XL probe is most relevant (eg, the proportion of XL probe use according to liver disease cause in table 2). The diagnostic performance in patients with NAFLD undergoing bariatric surgery versus patients not undergoing bariatric surgery and in those with a BMI more than or less than 30 kg/m² is shown in the appendix (p 4). The distribution of steatosis grades depended upon cohort type and diagnostic performance tended to be better in the bariatric cohorts, but differences are no longer relevant when distinguishing by BMI, although optimal cutoffs differed by subcohort. The correct probe choice in the NAFLD cohort is a matter of some debate, and the penetration depth of even the XL probe can be a limiting factor in patients who are severely obese.[13,42] The diagnostic performance using the manufacturer's recommendation for probe choice gave similar results—eg, for the NAFLD cohort, optimal cutoffs were within 1 dB/m of the optimal cutoffs calculated here (appendix pp 4–5).

If the sensitivity was set at 0·9 instead of the Youden cutoff approach, then the optimal cutoff in patients with NAFLD was 263 dB/m (95% CI 256–270) for S0 versus S1 to S3, 286 dB/m (282–292) for S0 to S1 versus S2 to S3, and 297 dB/m (286–307) for S0 to S2 versus S3. The values for specificity at these optimal cutoffs are 0·500 (95% CI 0·396–0·594) for S0 versus S1 to S3, 0·394 (0·349–0·438) for S0 to S1 versus S2, and 0·344 (0·313–0·375) for S0 to S2 versus S3. If the specificity is set at 0·9, optimal cutoff values are 354 dB/m (95% CI 316–373) for S0 versus S1 to S3, 372 dB/m (316–373) for S0 to S1 versus S2 to S3, and 385 dB/m (377–390) for S0 to S2 versus S3 with commensurate sensitivities of 0·520 (95% CI 0·491–0·548) for S0 versus S1 to S3, 0·255 (0·225–0·286) for S0 to S1 versus S2 to S3, and 0·224 (0·185–0·265) for S0 to S2 versus S3.

The diagnostic performance in patients with NAFLD with use of two sets of established cutoffs with and without adjustment of CAP values for liver disease

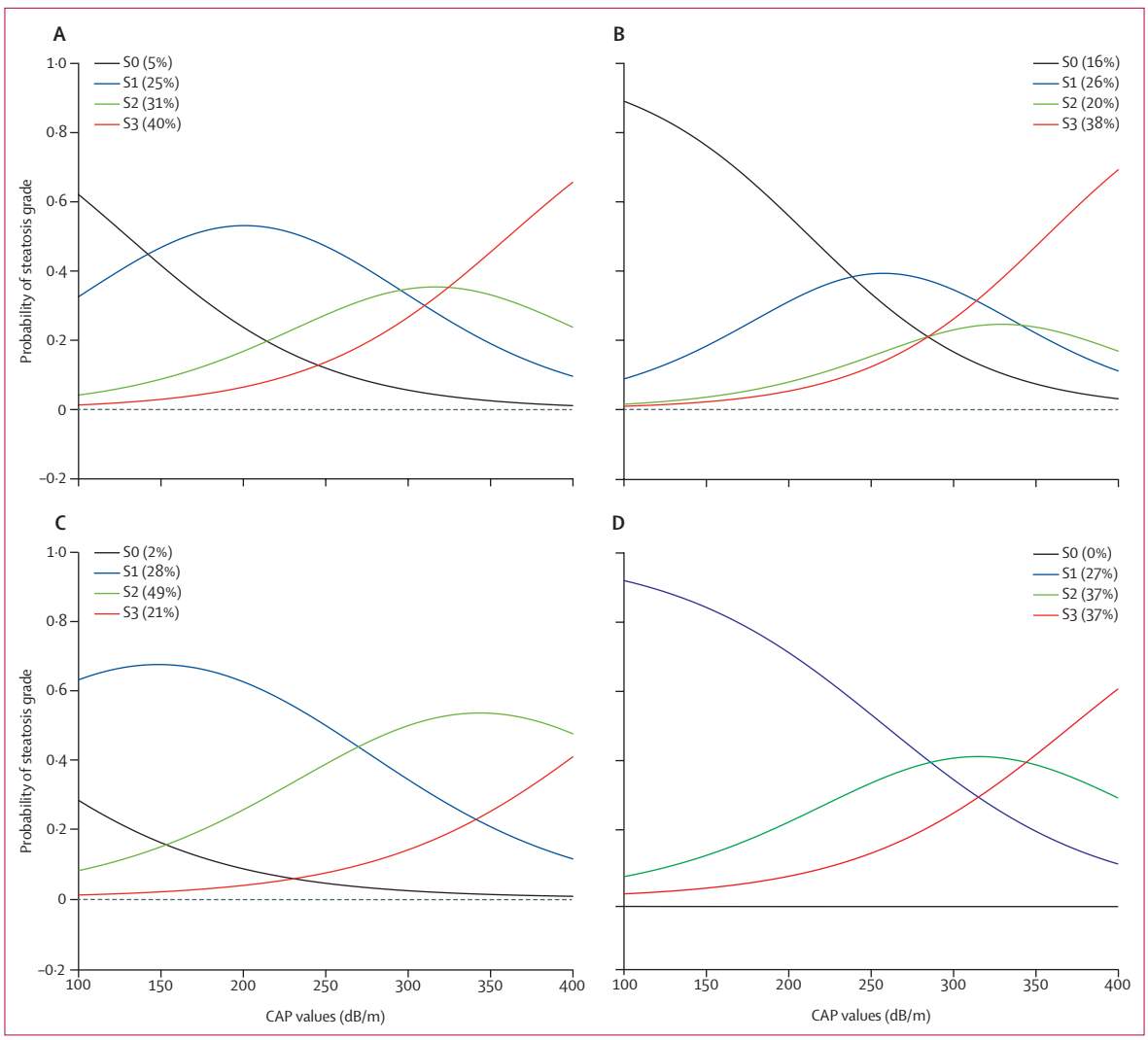

Figure 2: **Probability of steatosis grade as a function of CAP for patients with non-alcoholic fatty liver disease**
The prevalence of each grade (S0 to S3) is shown in parentheses. The choice of prevalence is based on the observed values from Eddowes et al, 2019[14] (A), Naveau et al, 2017[23] (B), Chan et al, 2018[26] (C), and Baumeler et al, 2019[15] (D). CAP=controlled attenuation parameter. Steatosis was defined according to the number of affected hepatocytes: S0 (<5%), S1 (5–33%), S2 (34–66%), and S3 (>66%).

cause, diabetes, and BMI according to previous studies derived from M probe data[11,14] is shown in the appendix (p 5). The cutoffs taken from our previous individual patient data meta-analysis[11] with mixed liver disease aetiologies and the M probe resulted in sensitivities greater than 0·9, but very poor specificities ranging from 0·25 to 0·42. Use of NAFLD derived cutoffs did not lead to satisfactory sensitivity and specificity, ranging from 0·57 to 0·74. Adjusting CAP values, but retaining the individual patient data meta-analysis cutoffs,[11] led to sensitivity and specificity roughly similar to those found with NAFLD derived cutoffs.

We estimated the steatosis grade probability at a given CAP value using ordinal regression (figure 2). Such probabilities are closely related to the positive predictive value and negative predictive value and depend strongly on prevalence. In some prevalence scenarios taken from the papers in this analysis, there are CAP values for which three or even all four steatosis grades are probable and distinction between them is therefore poor.

Although individual patient data is used in this study, the meta-analysis cannot be treated as a single study. To assess potential study effects, we derived AUCs for each study in this individual patient data meta-analysis (figure 3). The numbers were recalculated using only the patients included in this individual patient data meta-analysis and a standardised optimisation Youden method. Studies varied substantially and increasing NAFLD prevalence relative to other liver diseases tended to lead to poorer diagnostic performance.

The studies included in this individual patient data meta-analysis reported automated choice of probes for

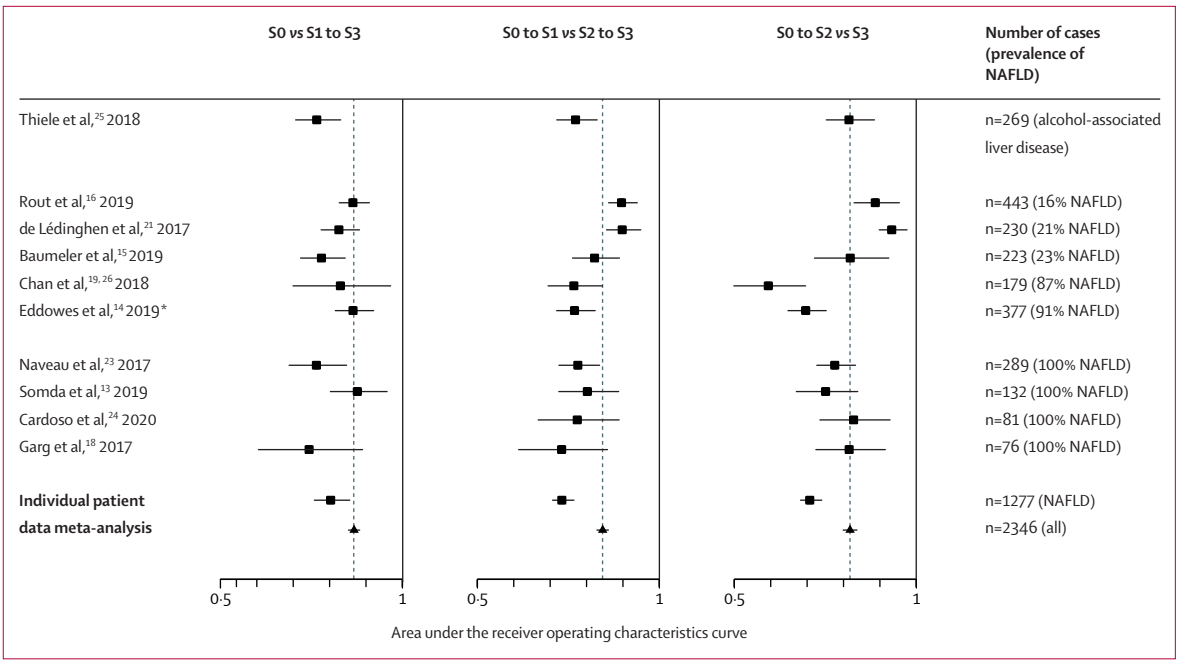

Figure 3: **Area under the receiver operating characteristics curve by data source**
Data source arranged by increasing prevalence of NAFLD. The areas under the curve from the current analysis and the papers that comprise its data are ordered by aetiology. Lai et al[19] only contributed 26 patients not in Chan et al.[19,26] Cardoso et al[24] had no S0 patients. NAFLD=non-alcoholic fatty liver disease. *In Eddowes et al,[14] only patients with suspected NAFLD were included. In 30 of 47 patients with S0, another diagnosis was established.

895 patients (four studies), use of both probes for 511 patients (five studies), choice of probe based on BMI for 443 patients (one study), choice based on SLD for 421 patients (two studies), and only used the XL probe for 76 patients (one study). However, in some cases, if the selected probe failed, then the other one was used. The XL probe was correctly chosen according to our definitions in 1050 of 2346 patients and the M probe in 1289 patients of 2346 including 20 cases with a BMI greater than 35 kg/m² but appropriate SLD. For seven patients, the correct probe choice could not be established as BMI was unavailable.

930 (89%) of 1050 correct XL measurements were made on patients with NAFLD. In contrast to the previous individual patient data meta-analysis,[11] steatosis grade S0 was dominated by liver diseases other than NAFLD (654 [87%] of 750) whereas all other liver disease aetiologies were under-represented relative to NAFLD for S1 to S3 (415 [26%] of 1596; appendix p 6). Notably, 75% of the NAFLD cohort with S0 (72 of 96 patients) were patients who had undergone bariatric surgery.

CAP values from both probes were available in 527 patients (307 [58%] with NAFLD). A Bland-Altman plot (appendix p 7) showed that there is little bias (4·4 dB/m) and essentially no dependence on the mean CAP value over a wide range of BMI values. However, the typical discrepancy between the M and XL probe in a given patient is large (mean absolute difference of 30 dB/m and a 95% predictive interval of –78 to 82 dB/m at a mean CAP value of 250 dB/m).

For all patients, the correct probe (M or XL) was also included as a covariate in a linear mixed model in which study was the random term. This multivariate analysis showed that compared with patients without steatosis, CAP values for patients with steatosis differed by about 30 dB/m for those with S1, 62 dB/m for S2, and 81 dB/m for S3 (table 3). Liver disease aetiology, BMI, sex, aspartate aminotransferase, and diabetes also affected CAP estimates, but the effect of probe choice was small. The SD of the random term was 10·3 dB/m with a residual of 46·5 dB/m and ANOVA indicated that it differed significantly from a linear model without the random term ($p<0·0001$), indicating that differences between studies are not negligible despite inclusion of covariates. However, estimates were not substantially changed in a linear model with study as a fixed effect despite accounting for steatosis grades, and although the study term was significant as a whole in this linear model ($p<0·0001$), no single study differed significantly from the arbitrarily chosen reference study. Similar results were found in patients with NAFLD alone, except that aspartate aminotransferase was no longer identified as a relevant or significant covariate (appendix p 8). Adjusting the CAP value in patients with NAFLD by 10 dB/m for those with diabetes and for men by 2·6 dB/m per BMI point more than or less than 25 kg/m² up to a maximum of 50 kg/m² changed the AUC from 0·807 to 0·826 (S0 *vs* S1 to S3), 0·736 to 0·762 (S0 to S1 *vs* S2 to S3), and 0·711 to 0·702 (S0 to S2 *vs* S3).

For some patients, the M and XL probes might both be viable choices—ie, when penetration depth is close to

the crossover from the M to the XL probe. Unfortunately, SLD was unavailable in datasets for which both probes were used. Hence, we considered patients with NAFLD with BMI greater than 23 kg/m² and less than 30 kg/m², in which more than half (75 [56%] of 135) had a discrepancy in steatosis grade when steatosis categories were allocated with the cutoffs from table 2. In 27 (20%) of 135 cases there was a difference of at least two steatosis grades (appendix p 9). The M probe yielded the correct result in 33 (24%) of 135 cases and the XL probe in 37 (27%) of 135 cases.

Diagnostic properties could be affected by unreliable measurements. With a threshold IQR of more than 40 dB/m, 32% of measurements were defined as unreliable in the NAFLD cohort. However, the correlation between CAP and percentage of hepatocytes does not improve if these unreliable measurements are excluded, changing from r=0·452 to r=0·449 with the exclusion of unreliable measurements (appendix p 10). With a threshold based on IQR to median ratio, 3% of measurements were deemed unreliable at a ratio cutoff at 0·3, 11% at 0·2, and 46% at 0·1. The correlations between CAP and proportion of steatotic hepatocytes worsened with more stringent IQR to median ratio cutoffs (r=0·436 for 0·3, 0·399 for 0·2, and 0·337 for 0·1, appendix p 10). Hence, we were not able to find evidence that any reliability criterion cutoff value improved performance.

Discussion

This comprehensive individual patient data meta-analysis on the diagnostic capabilities of CAP, including the XL probe for non-invasive grading of hepatic fat, found that diagnostic properties depended substantially on cause of hepatic steatosis. Higher grades of steatosis in patients with NAFLD could not be distinguished reliably for individual patients and the areas under the curve for the XL probe were 0·819 (95% CI 0·769–0·869) with optimal cutoff at 297 dB/m for distinguishing S0 from S1 to S3, 0·754 (0·720–0·787) with cutoff at 317 dB/m for distinguishing S0 to S1 from S2 to S3, and 0·717 (0·684–0·751) with cutoff at 333 dB/m for distinguishing S0 to S2 from S3. Mean CAP values from the available measurements using both M and XL probes in the same patient were similar, although the mean absolute difference between them was high, resulting in misclassification of 50% of NAFLD cases (appendix p 7). Diagnostic accuracy did not improve by applying proposed reliability criteria.[36,37] After adjusting for steatosis grade, we verified that CAP depended significantly on liver disease cause, diabetes, and BMI and found moreover that CAP for men is about 12 dB/m higher than for women. Unexpectedly, diagnostic performance in patients undergoing bariatric surgery was somewhat better, which might in part be due to highly standardised recruitment and procedures and large liver biopsy samples.

	Estimate (95% CI), dB/m
Intercept	239 (219 to 259)
Steatosis (S0 reference)	
S1	30·4 (24·2 to 36·8)
S2	62·1 (55·1 to 69·4)
S3	81·0 (73·6 to 88·5)
Liver disease aetiology† (alcohol-associated liver disease reference)	
Hepatitis B	−17·2 (−34·8 to 0·2)
Hepatitis C	−18·2 (−35·6 to −0·8)
NAFLD or NASH	0·6 (−16·2 to 17·4)
Other	−20·6 (−37·7 to −3·5)
XL probe (M probe reference)	6·5 (−0·5 to 12·6)
BMI (per 1 kg/m² increase compared with 25 kg/m² reference)	2·57 (2·11 to 2·97)
Type 2 diabetes (presence)	13·6 (8·7 to 18·8)
Sex (male *vs* female reference)	12·0 (7·7 to 16·4)
Alanine aminotransferase (per doubling of value)	1·06 (−2·07 to 4·34)
Aspartate aminotransferase (per doubling of value)	−3·81 (−7·46 to −0·29)

Data are estimate (95% CI). Data are shown from the fixed terms in a mixed model with CAP value as the dependent variable and study as a random term. The estimates describe how each variable is associated with the CAP result, even after accounting for the others. Each variable estimate is derived from a model containing all other variables as covariates. For example, patients with type 2 diabetes will have CAP values a mean of 13·6 dB/m higher than patients without type 2 diabetes, after accounting for liver disease cause, steatosis grade, BMI, sex, alanine aminotransferase, and aspartate aminotransferase. Steatosis was defined according to the number of affected hepatocytes: S0 (<5%), S1 (5–33%), S2 (34–66%), and S3 (>66%). BMI=body-mass index. CAP=controlled attenuation parameter. NALFD=non-alcoholic fatty liver disease. NASH=non-alcoholic steatohepatitis.

Table 3: **Covariates affecting CAP values**

Although our quantitative results agree on the whole with the published literature (compare with figure 3), the picture that emerges when the results are viewed in this pooled fashion is new, and the conclusions we draw are different. Individual studies have suggested that the diagnosis of steatosis in NAFLD with CAP works well with the XL probe in patients who are obese—ie, that the AUC for S0 versus S1 to S3 is high. However, some of these studies contain cohorts with mixed liver disease aetiologies in which S0 is dominated by patients without NAFLD.[15,16,21] Others are pure NAFLD cohorts, but with a small number of patients with S0.[13,14,18,23,26] The definition of NAFLD formally excludes S0, so the appearance of these patients in these cohorts merits further consideration. In our data, patients who were S0 came either from studies with bariatric surgery cohorts,[13,18,23] in which histology is routinely obtained during surgery and irrespective of liver disease severity, or they come from a screening setting in which most patients who were S0 were ultimately classified as non-NAFLD.[14] Notably, the patients with S0 and NAFLD did not have advanced fibrosis, suggesting that their lack of steatosis was not the result of hepatocyte fat loss in late-stage fibrotic NASH. Because the S0 group is small and atypical, the

important clinical question of whether steatosis can be detected by CAP in a screening setting for individuals at risk for NAFLD cannot be answered adequately with available data.

Our group did a previous individual patient data meta-analysis on M probe CAP and found results supportive of CAP, with AUCs ranging from 0·82 to 0·88.[11] The data were restricted to individuals with lower BMIs (BMI >35 kg/m² was an exclusion criterion) with M probe measurements and were dominated by viral hepatitis. Moreover, there were almost no patients who were S0 and classified as having NAFLD and diagnostic properties for this aetiology alone could not be provided. An early conventional meta-analysis did investigate diagnostic properties in NAFLD, but looked mainly at M probe data despite many patients being obese and concluded that diagnostic capabilities should be viewed with caution.[43] Only two of the nine studies included in that previous meta-analysis used the XL probe and thus qualified for the current analysis and only one could be included due to data sharing issues. Our current data suggest the need to be cautious with CAP for steatosis grading in NAFLD even if the XL probe is used (table 2).

In the current meta-analysis, patients with viral hepatitis tended to have higher BMIs than did the patients with viral hepatitis in our original individual patient data meta-analysis[11] because of the availability of the XL probe. The low prevalence of high-grade steatosis and small total numbers meant that diagnostic performance could not be analysed in great depth for this aetiology. AUC was slightly lower here for S0 versus S1 to S3 than the previous individual patient data meta-analysis (0·769 *vs* 0·82), whereas it was similar for S0 to S1 versus S2 to S3 (0·847 *vs* 0·86). Patients with alcohol-associated liver disease in the current analysis had a fairly high prevalence of obesity (20%) despite infrequent use of the XL probe (8%), but they derived almost exclusively from a single study,[25] suggesting that more data for this aetiology are needed.

There has been some debate about optimal cutoffs and how they might depend on liver disease cause, probe (M *vs* XL), and anthropometry.[11,14,26,44,45] Considering the different penetration depths of the probes, it is essential to distinguish two types of comparison—between two probes used in a given patient in whom such data are available, in which case small mean differences were found of less than 10 dB/m,[26,44,46] and between probes each used in the appropriate cohort (majority of data), in which case large differences in cutoffs are found unless models correct for anthropometry and comorbidities. Here, we can confirm both observations but conclude that the small differences between the probes is not of clinical relevance if the correct probe is used. The choice of cutoff then might benefit from adjustments for liver disease cause, diabetes, and BMI, as recommended in our previous individual patient data meta-analysis,[11] which for NAFLD in the current paper would lead to a shift for example from 280 dB/m (S0 to S2 *vs* S3) to about 330 dB/m, which is close to the cutoff of 337 dB/m identified by Eddowes and colleagues.[14] Without such adjustments, sensitivity is very high, but specificity is very low. Other authors have already noted that adjustments do not necessarily improve AUC or other markers of overall diagnostic performance.[26,37] Adjustments do however more equally balance sensitivity and specificity, which could be relevant if both ruling in and ruling out steatosis are of interest. When only one of the two is the diagnostic goal, then it is typical to optimise based on either sensitivity or specificity of, for example, 0·9, as demonstrated by Eddowes and colleagues.[14] A specificity of 0·9 can lead to very high cutoffs, which are no longer of clinical use. In the context of NAFLD, the current analysis suggested that adjustments might have the potential to improve AUC slightly, but this observation requires validation.

It is interesting to speculate as to what physiological mechanism could affect CAP measurements in patients who have diabetes after taking BMI, sex, and other covariates into account. Evidence suggests that micro-vesicular steatosis is related to progression of NAFLD,[47] and might be more prevalent among patients with diabetes.[48] Because small fat droplets lead to more signal attenuation, this mechanism might lead to higher CAP values. Furthermore, patients with diabetes might have a greater prevalence of steatohepatitis, which could also affect CAP. Such considerations could also be relevant in explaining aetiology specific differences. Beyond physiological mechanisms, the very different prevalence of steatosis depending on liver disease cause leads to different optimal cutoffs in ROC analyses.

The non-invasive quantification of hepatic fat requires critical appraisal as its clinical implications are a matter of ongoing debate.[49] Although fat accumulation in hepatocytes is considered a key mechanism in the natural history of fatty liver disease, long-term observational studies have not found that the histological degree of steatosis correlates with clinical endpoints such as hepatic related mortality.[50] This finding challenges the need for estimating hepatic fat, but associations have been found between liver fat and progression of fibrosis[51] and between liver fat and metabolic syndrome, even after adjusting for NASH.[52] Moreover, current outcome data are biased due to the need for an invasive reference standard and fairly small sample sizes compared with, for example, blood pressure outcome studies in cardiology. Liver fat assessed by histology refers to a proportion of affected cells by surface area; values derived from MRI and CAP still require studies with long-term follow-up and linkage to outcomes, as steatosis determined by these methods might be associated with outcomes where histological steatosis is not.[9] In addition, the value of serial liver fat quantification remains to be defined for determining prognosis and treatment efficacy.[7]

Another potential clinical benefit for CAP can be in identifying the small proportion of individuals suspected of having NAFLD with S0 who ultimately turn out to have alternative diagnoses.[14] Studies focusing on screening cohorts cannot use liver biopsies as a reference standard, but could use MRI, which is very sensitive to hepatic fat.[8] Up to now, only a few studies have compared both MRI and CAP with histology with a modest sample size[20,53,54] or without patients with S0,[55] but all of them concluded that MRI outperforms CAP. Such studies can also address the value of continuous tests for steatosis severity, which is not possible with traditional steatosis grading. However, the clinical availability and costs of MRI imply that combinations of screening tools need to be considered.[9]

The data presented here suggest that CAP could be of epidemiological value but raise questions about what role it could play in the general practitioner's toolbox. Previous analyses have shown good diagnostic performance in patients with viral hepatitis using the M probe,[11] and the current analysis suggests only a slight deterioration when extending the population to patients with a higher BMI and use of the XL probe. However, it is of limited diagnostic value for grading steatosis in patients who are obese with NAFLD, which emphasises the inherent technical limitations of this ultrasound-based steatosis quantification approach. Using CAP, ultrasound signal attenuation is not exclusively induced by hepatocellular lipid droplets but is also related to subcutaneous tissue and abdominal wall properties,[56] which become especially relevant in patients with metabolic risk factors and might explain our observations.

Despite such considerations, CAP still represents the first highly standardised and evaluated approach using ultrasound signal modulation for steatosis quantification. Technical modifications of CAP, such as special algorithms for patients who are severely obese (BMI >40 kg/m²),[13] could improve diagnostic precision. In addition, further developments by other manufacturers use two-dimensional B-mode image control of the target region, which in pilot studies enhances performance and merits further evaluation.[57,58]

Strengths of our analysis are that it uses the largest comprehensive database on biopsy-controlled CAP measurements and that the mean time interval between the index (CAP) and reference tests (biopsy) was very short. The data were derived from many expert international groups, with only small differences between studies. The large number of paired XL and M probe measurements in patients who were overweight without a priori indication for either probe facilitated important comparisons between the two. This pooling of the available patients with NAFLD showed that a considerable proportion of this population derived from bariatric surgery cohorts, but not from a typical screening target population. More data are required from typical screening cohorts.

A general weakness of studies with invasive reference standards is the so-called verification bias, as described by O'Sullivan and colleagues.[59] It is incorrect to believe that sensitivity and specificity estimated from a complete case analysis are unbiased compared with a population without biopsy, which again underscores the need for CAP screening studies with MRI as the reference standard. The high prevalence of significant fibrosis illustrated that indication for biopsy was based on the suspicion of advanced liver disease in many of the patients. Moreover, patients with bariatric surgery are recruited with less emphasis on liver disease and tended to have fewer so-called grey zone patients with S1 and S2 so that diagnostic performance in these individuals was better. This finding might also hold for screening populations of interest without indication for liver biopsy. A related issue is the spectrum effect, describing a dependence on sensitivity and specificity with regard to patient characteristics.[60] This dependence can induce spectrum bias if populations are selected on the basis of such characteristics and results are generalised to a different target population, which might explain in part the non-negligible differences between studies. The QUADAS-2 assessment showed that there is some potential for bias because of patient selection and imprecise use of the index test. These biases are expected to be small compared with the verification and spectrum biases, however, when CAP is used in a screening population. A specific limitation of this individual patient data meta-analysis is that we could not include data from the US NASH Clinical Research Network because of National Institutes of Health restrictions.[17,20,27] However, these studies in total would presumably have supplied data from about 300 patients to our analysis (appendix p 3) and only about 20 with S0. Given the somewhat worse diagnostic performance of CAP reported in table 4 of Siddiqui et al,[27] this absence of data could be expected to bias our results slightly toward optimistic estimates. It is not feasible in an intercontinental diagnostic individual patient data meta-analysis to obtain central histological readings. However, as steatosis is known to have good inter-observer and intra-observer agreement, in contrast to other NASH parameters,[32,61] this is likely only to have led to a small bias, if any. The choice of correct probe in the studies analysed here did not always follow the manufacturer's recommendations and the effect of this discrepancy on performance would be interesting, but was not possible to determine in this study given the nature of the data. Future studies should firmly adhere to recommendations based on current guidelines and the manufacturer's recommendations. Finally, the strict definition of NAFLD precluding concomitant diseases could be inherently problematic as it neglects metabolically driven processes in liver disease of other aetiologies. A proposal for a revision of the definition of metabolic liver disease addresses this issue but was not

available in the current data set.[62] Such considerations could also be relevant for patients with viral hepatitis and concomitant fatty liver disease.

This analysis completes a decade of CAP research for the grading of steatosis, but does not focus on advancements from 2019–20 for the detection of steatohepatitis, which combine CAP with liver stiffness measurement and aspartate aminotransferase levels and might be useful for guiding medical interventions in the future.[63] However, the imprecision of CAP-based steatosis grading, especially when important covariates are left unconsidered, sheds light on the capabilities of non-invasive NASH detection and highlights the necessity for independent studies with various designs.

In conclusion, CAP cutoffs vary according to liver disease cause, and effectively recognise significant steatosis in patients with viral hepatitis. CAP values alone relative to rigid cutoffs provide little clinically valuable information for grading steatosis in patients with NAFLD despite use of the XL probe. On the basis of a subset of the data, the difference between M and XL probe values in a given patient might be large, but their mean difference in a given population is quite small. Current data for S0 in the context of NAFLD are scant and heavily influenced by data from bariatric surgery cohorts, so that firm conclusions on diagnostic performance cannot yet be drawn. Knowledge of steatosis prevalence and covariates might help interpret CAP use and application in screening scenarios requires further research beyond the traditional histological reference standard.

Contributors

DP, VB, JW, and TK were responsible for study concept and design, literature search, collecting data from participating centres, data review, verification and analysis, interpretation of data, drafting of the manuscript, statistical analysis, and obtaining funding. The remaining authors were responsible for drafting of the protocol, obtaining original data, coding and providing data, interpretation of data, and critical revision of the manuscript.

Declaration of interests

DP, TK, and JW received unrestricted research grants from Echosens, and TK participated in a clinical advisory board meeting for Echosens. MS was an Echosens scientific employee in the research and development department, but contributed in her role as scientist and author of papers in this meta-analysis. Echosens provided the FibroScan device to Antoine-Béclère Hospital, but the authors CSV, GP, and SN did not receive funding from the manufacturers for their research. VW served as a consultant and speaker for Echosens. All other authors declare no competing interests.

Data sharing

Because this is an individual patient data meta-analysis, data cannot be provided without the consent of all parties involved. Individual datasets should be requested from the authors of the relevant papers.

Acknowledgments

This project was supported in part by the German Federal Ministry of Education and Research (BMBF, FKZ 01EO1001). Echosens provided funding but did not influence study design and were not involved in data analysis. Data collection at Odense University Hospital received support from Innovation Fund Denmark, the Challenge Grant MicrobLiver (grant number NNF15OC0016692) from the Novo Nordisk Foundation, and EU's Horizon 2020 research and innovation programme for GALAXY (grant agreement number 668031). Data collection at the Federal University of Rio de Janeiro received support from the National Council for Research and Technology (CNPq, Brazil). QMA is a Newcastle National Institute of Health Research (NIHR) Biomedical Research Centre investigator. PNN was supported by the NIHR Birmingham Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the UK Department of Health.

References

1 Castera L, Friedrich-Rust M, Loomba R. Noninvasive assessment of liver disease in patients with nonalcoholic fatty liver disease. *Gastroenterology* 2019; **156:** 1264–81.

2 European Association for Study of Liver. EASL-ALEH Clinical Practice Guidelines: non-invasive tests for evaluation of liver disease severity and prognosis. *J Hepatol* 2015; **63:** 237–64.

3 Angulo P, Kleiner DE, Dam-Larsen S, et al. Liver Fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology* 2015; **149:** 389–97.

4 Ekstedt M, Hagström H, Nasr P, et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. *Hepatology* 2015; **61:** 1547–54.

5 McPherson S, Hardy T, Henderson E, Burt AD, Day CP, Anstee QM. Evidence of NAFLD progression from steatosis to fibrosing-steatohepatitis using paired biopsies: implications for prognosis and clinical management. *J Hepatol* 2015; **62:** 1148–55.

6 Liu K, Wong VW-S, Lau K, et al. Prognostic value of controlled attenuation parameter by transient elastography. *Am J Gastroenterol* 2017; **112:** 1812–23.

7 Wong VW-S. Predicting NASH response with liver fat: are we back to square one? *J Hepatol* 2020; **72:** 386–88.

8 Caussy C, Reeder SB, Sirlin CB, Loomba R. Noninvasive, quantitative assessment of liver fat by MRI-PDFF as an endpoint in NASH trials. *Hepatology* 2018; **68:** 763–72.

9 Karlas T, Petroff D, Wiegand J. Collaboration, not competition: the role of magnetic resonance, transient elastography, and liver biopsy in the diagnosis of nonalcoholic fatty liver disease. *Gastroenterology* 2017; **152:** 479–81.

10 Sasso M, Beaugrand M, de Ledinghen V, et al. Controlled attenuation parameter (CAP): a novel VCTE guided ultrasonic attenuation measurement for the evaluation of hepatic steatosis: preliminary study and validation in a cohort of patients with chronic liver disease from various causes. *Ultrasound Med Biol* 2010; **36:** 1825–35.

11 Karlas T, Petroff D, Sasso M, et al. Individual patient data meta-analysis of controlled attenuation parameter (CAP) technology for assessing steatosis. *J Hepatol* 2017; **66:** 1022–30.

12 Sasso M, Audière S, Kemgang A, et al. Liver steatosis assessed by controlled attenuation parameter (CAP) measured with the XL probe of the FibroScan: a pilot study assessing diagnostic accuracy. *Ultrasound Med Biol* 2016; **42:** 92–103.

13 Somda S, Lebrun A, Tranchart H, et al. Adaptation of controlled attenuation parameter (CAP) measurement depth in morbidly obese patients addressed for bariatric surgery. *PLoS One* 2019; **14:** e0217093.

14 Eddowes PJ, Sasso M, Allison M, et al. Accuracy of FibroScan controlled attenuation parameter and liver stiffness measurement in assessing steatosis and fibrosis in patients with nonalcoholic fatty liver disease. *Gastroenterology* 2019; **156:** 1717–30.

15 Baumeler S, Jochum W, Neuweiler J, Bergamin I, Semela D. Controlled attenuation parameter for the assessment of liver steatosis in comparison with liver histology: a single-centre real life experience. *Swiss Med Wkly* 2019; **149:** w20077.

16 Rout G, Kedia S, Nayak B, et al. Controlled attenuation parameter for assessment of hepatic steatosis in Indian patients. *J Clin Exp Hepatol* 2019; **9:** 13–21.

17 Vuppalanchi R, Siddiqui MS, Van Natta ML, et al. Performance characteristics of vibration-controlled transient elastography for evaluation of nonalcoholic fatty liver disease. *Hepatology* 2018; **67:** 134–44.

18 Garg H, Aggarwal S, Shalimar, et al. Utility of transient elastography (fibroscan) and impact of bariatric surgery on nonalcoholic fatty liver disease (NAFLD) in morbidly obese patients. *Surg Obes Relat Dis* 2018; **14:** 81–91.

19 Lai L-L, Wan Yusoff WNI, Vethakkan SR, Nik Mustapha NR, Mahadeva S, Chan W-K. Screening for non-alcoholic fatty liver disease in patients with type 2 diabetes mellitus using transient elastography. *J Gastroenterol Hepatol* 2019; **34:** 1396–403.

20 Park CC, Nguyen P, Hernandez C, et al. Magnetic resonance elastography vs transient elastography in detection of fibrosis and noninvasive measurement of steatosis in patients with biopsy-proven nonalcoholic fatty liver disease. *Gastroenterology* 2017; **152:** 598–607.

21 de Lédinghen V, Hiriart J-B, Vergniol J, Merrouche W, Bedossa P, Paradis V. Controlled attenuation parameter (CAP) with the XL probe of the Fibroscan: a comparative study with the M Probe and liver biopsy. *Dig Dis Sci* 2017: **62:** 2569–77.

22 Chan W-K, Nik Mustapha NR, Wong GL-H, Wong VW-S, Mahadeva S. Controlled attenuation parameter using the FibroScan XL probe for quantification of hepatic steatosis for non-alcoholic fatty liver disease in an Asian population. *United European Gastroenterol J* 2017; **5:** 76–85.

23 Naveau S, Voican CS, Lebrun A, et al. Controlled attenuation parameter for diagnosing steatosis in bariatric surgery candidates with suspected nonalcoholic fatty liver disease. *Eur J Gastroenterol Hepatol* 2017; **29:** 1022–30.

24 Cardoso AC, Cravo C, Calçado FL, et al. The performance of M and XL probes of FibroScan for the diagnosis of steatosis and fibrosis on a Brazilian nonalcoholic fatty liver disease cohort. *Eur J Gastroenterol Hepatol* 2020; **32:** 231–38.

25 Thiele M, Rausch V, Fluhr G, et al. Controlled attenuation parameter and alcoholic hepatic steatosis: Diagnostic accuracy and role of alcohol detoxification. *J Hepatol* 2018; **68:** 1025–32.

26 Chan W-K, Mustapha NR, Mahadeva S, Wong VW-S, Cheng JY-K, Wong GL-H. Can the same controlled attenuation parameter cut-offs be used for M and XL probes for diagnosing hepatic steatosis? *J Gastroenterol Hepatol* 2018; **33:** 1787–94.

27 Siddiqui MS, Vuppalanchi R, van Natta ML, et al. Vibration-controlled transient elastography to assess fibrosis and steatosis in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2019; **17:** 156–63.

28 Karlas T, Beer S, Babel J, et al. Do we need controlled attenuation parameter adjustment for fibrosis estimation in non-alcoholic fatty liver disease patients? *Hepatology* 2017; **65:** 2126–28.

29 Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005; **41:** 1313–21.

30 Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155:** 529–36.

31 Bedossa P. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. *Hepatology* 1994; **20:** 15–20.

32 Merriman RB, Ferrell LD, Patti MG, et al. Correlation of paired liver biopsies in morbidly obese patients with suspected nonalcoholic fatty liver disease. *Hepatology* 2006; **44:** 874–80.

33 Dietrich CF, Bamber J, Berzigotti A, et al. EFSUMB guidelines and recommendations on the clinical use of liver ultrasound elastography, Update 2017 (long version). *Ultraschall Med* 2017; **38:** e16–47.

34 Shen F, Zheng R-D, Shi J-P, et al. Impact of skin capsular distance on the performance of controlled attenuation parameter in patients with chronic liver disease. *Liver Int* 2015; **35:** 2392–400.

35 Boursier J, Zarski J-P, de Ledinghen V, et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. *Hepatology* 2013; **57:** 1182–91.

36 Wong VW-S, Petta S, Hiriart J-B, et al. Validity criteria for the diagnosis of fatty liver by M probe-based controlled attenuation parameter. *J Hepatol* 2017; **67:** 577–84.

37 Semmler G, Wöran K, Scheiner B, et al. Novel reliability criteria for controlled attenuation parameter assessments for non-invasive evaluation of hepatic steatosis. *United European Gastroenterol J* 2020; **8:** 321–31.

38 López-Ratón M, Rodríguez-Álvarez MX, Suárez CC, Sampedro FG. OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *J Stat Softw* 2014; **61:** 1–36.

39 Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 1927; **22:** 209–12.

40 Venables WN, Ripley BD. Modern applied statistics with S. 4th ed. New York: Springer; 2002.

41 Thiele M, Detlefsen S, Sevelsted Møller L, et al. Transient and 2-dimensional shear-wave elastography provide comparable assessment of alcoholic liver fibrosis and cirrhosis. *Gastroenterology* 2016; **150:** 123–33.

42 Karlas T, Dietrich A, Peter V, et al. Evaluation of transient elastography, acoustic radiation force impulse imaging (ARFI), and enhanced liver function (ELF) score for detection of fibrosis in morbidly obese patients. *PLoS One* 2015; **10:** e0141649.

43 Pu K, Wang Y, Bai S, et al. Diagnostic accuracy of controlled attenuation parameter (CAP) as a non-invasive test for steatosis in suspected non-alcoholic fatty liver disease: a systematic review and meta-analysis. *BMC Gastroenterol* 2019; **19:** 51.

44 Caussy C, Brissot J, Singh S, et al. Prospective, same-day, direct comparison of controlled attenuation parameter with the m vs the xl probe in patients with nonalcoholic fatty liver disease, using magnetic resonance imaging-proton density fat fraction as the standard. *Clin Gastroenterol Hepatol* 2020; **18:** 1842–50.

45 Shalimar KR, Kumar R, Rout G, et al. Body mass index-based controlled attenuation parameter cut-offs for assessment of hepatic steatosis in non-alcoholic fatty liver disease. *Indian J Gastroenterol* 2020; **39:** 32–41.

46 Oeda S, Takahashi H, Imajo K, et al. Accuracy of liver stiffness measurement and controlled attenuation parameter using FibroScan M/XL probes to diagnose liver fibrosis and steatosis in patients with nonalcoholic fatty liver disease: a multicenter prospective study. *J Gastroenterol* 2020; **55:** 428–40.

47 Tandra S, Yeh MM, Brunt EM, et al. Presence and significance of microvesicular steatosis in nonalcoholic fatty liver disease. *J Hepatol* 2011; **55:** 654–59.

48 Younossi ZM, Golabi P, de Avila L, et al. The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: a systematic review and meta-analysis. *J Hepatol* 2019; **71:** 793–801.

49 Ballestri S, Nascimbeni F, Lugari S, Lonardo A, Francica G. A critical appraisal of the use of ultrasound in hepatic steatosis. *Expert Rev Gastroenterol Hepatol* 2019; **13:** 667–81.

50 Angulo P, Bugianesi E, Bjornsson ES, et al. Simple noninvasive systems predict long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology* 2013; **145:** 782–89.

51 Ajmera V, Park CC, Caussy C, et al. Magnetic resonance imaging proton density fat fraction associates with progression of fibrosis in patients with non-alcoholic fatty liver disease. *Gastroenterology* 2018; **155:** 307–10.

52 Arulanandan A, Ang B, Bettencourt R, et al. Association between quantity of liver fat and cardiovascular risk in patients with nonalcoholic fatty liver disease independent of nonalcoholic steatohepatitis. *Clin Gastroenterol Hepatol* 2015; **13:** 1513–20.

53 Runge JH, Smits LP, Verheij J, et al. MR Spectroscopy-derived proton density fat fraction is superior to controlled attenuation parameter for detecting and grading hepatic steatosis. *Radiology* 2018; **286:** 547–56.

54 Imajo K, Kessoku T, Honda Y, et al. Magnetic resonance imaging more accurately classifies steatosis and fibrosis in patients with nonalcoholic fatty liver disease than transient elastography. *Gastroenterology* 2016; **150:** 626–37.

55 Harry S, Lai LL, Nik Mustapha NR, et al. Volumetric liver fat fraction determines grade of steatosis more accurately than controlled attenuation parameter in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2020; **18:** 945–53.

56 Patterson B, Miller DL. Experimental measurements of ultrasound attenuation in human chest wall and assessment of the mechanical index for lung ultrasound. *Ultrasound Med Biol* 2020; **46:** 1442–54.

57 Fujiwara Y, Kuroda H, Abe T, et al. The B-mode image-guided ultrasound attenuation parameter accurately detects hepatic steatosis in chronic liver disease. *Ultrasound Med Biol* 2018; **44:** 2223–32.

58 Ferraioli G, Maiocchi L, Raciti MV, et al. Detection of liver steatosis with a novel ultrasound-based technique: a pilot study using MRI-derived proton density fat fraction as the gold standard. *Clin Transl Gastroenterol* 2019; **10:** e00081.

59 O'Sullivan JW, Banerjee A, Heneghan C, Pluddemann A. Verification bias. *BMJ Evid Based Med* 2018; **23:** 54–55.

60 Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. *BMJ* 2016; **353:** i3139.

61 Pournik O, Alavian SM, Ghalichi L, et al. Inter-observer and intra-observer agreement in pathological evaluation of non-alcoholic fatty liver disease suspected liver biopsies. *Hepat Mon* 2014; **14:** e15167.

62 Eslam M, Newsome PN, Sarin SK, et al. A new definition for metabolic dysfunction-associated fatty liver disease: an international expert consensus statement. *J Hepatol* 2020; **73:** 202–09.

63 Newsome PN, Sasso M, Deeks JJ, et al. FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study. *Lancet Gastroenterol Hepatol* 2020; **5:** 362–73.