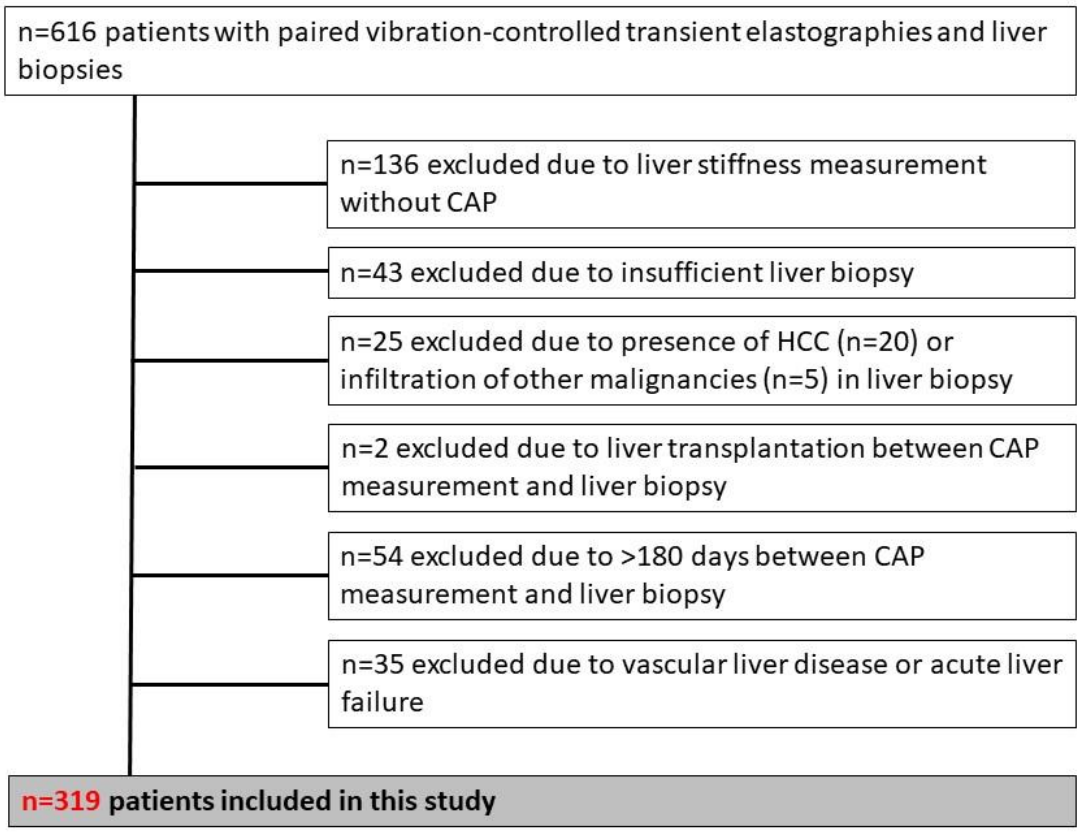
SUPPLEMENT

Supplementary Figure 1. Patient flow chart



Abbreviations: CAP – controlled attenuation parameter; IQR – interquartile range; HCC – hepatocellular carcinoma

Supplementary Table 1. Net reclassification improvement (NRI) for CAP corrections according to Karlas et al[13] for identifying patients with **(A)** any hepatic steatosis (HS≥S1), **(B)** HS≥S2, and **(C)** HS≥S3.

A: ≥S1: Karlas corrected vs. non-Karlas corrected

Event=any steatosis		Test 1 (CAP not corrected)		Total
Non-event		Abnormal=positive for steatosis	Normal	
Test 2 (CAP Karlas corrected)	Abnormal	162	2	164
		14	7	21
	Normal	15	47	62
		4	68	72
Total		177	49	226
		18	75	93

NRI_e = (2-15)/226 = -0.0575. NRI_{ne} = (4-7)/93 = -0.0322. NRI= -0.090.

B: ≥S2: Karlas corrected vs. non-Karlas corrected

Event=any steatosis		Test 1 (CAP not corrected)		Total
Non-event		Abnormal=positive for steatosis	Normal	
Test 2 (CAP Karlas corrected)	Abnormal	88	1	89
		51	2	53
	Normal	16	21	37
		13	127	140
Total		104	22	126
		64	129	193

NRI_e = (1-16)/126 = -0.127. NRI_{ne} = (13-2)/193 = 0.057. NRI= -0.070.

C: ≥S3: Karlas corrected vs. non-Karlas corrected

Event=any steatosis	Test 1 (CAP not corrected)	Total

Non-event		Abnormal=positive for steatosis	Normal	
Test 2 (CAP Karlas corrected)	Abnormal	37	1	
		66	0	
	Normal	12	9	
		25	169	
Total, split		49	10	59
		91	169	260

NRI_e = (1-12)/59 = -0.186. NRI_{ne} = (25-0)/260 = 0.096. NRI= -0.090.

Supplementary Table 2. Uni- and multivariate linear regression analyses investigating variables associated with **(A)** CAP values and **(B)** CAP IQR.

A	**Univariate**			**Multivariate (I)**		**Multivariate (II)**	
Patient characteristics	r	B	*P* value	B	*P* value	B	*P* value
Age, years	0.036	0.197	0.518	-	-	-	-
Male gender	0.047	7.096	0.402	-	-	-	-
BMI, kg x m^{-2}	0.181	0.665	**0.001**	0.231	0.139	0.270	0.080
Diabetes	0.329	57.748	**<0.001**	23.866	**0.002**	26.343	**0.001**
NAFLD	0.557	84.394	**<0.001**	53.939	**<0.001**	47.940	**<0.001**
Liver stiffness, kPa	-0.174	-0.727	**0.002**	0.223	0.244	-	-
Histological fibrosis stage	-0.190	-10.362	**0.001**	-	-	-2.763	0.259
Hepatic steatosis, %	0.556	1.421	**<0.001**	0.891	**<0.001**	0.887	**<0.001**
Bilirubin, mg/L	-0.088	-1.767	0.115	-	-	-	-
ALT, U/L	-0.020	-0.014	0.727	-	-	-	-
AP, U/L	-0.258	-0.204	**<0.001**	-0.066	0.083	-0.009	0.657
γ-GT, U/L	-0.155	-0.067	**0.006**	-0.015	0.463	-0.014	0.473

Multivariate (I+II): Due to multicollinearity, liver stiffness was included into model (I) and histological fibrosis stage into model (II).

B	**Univariate**			**Multivariate (I)**		**Multivariate (II)**	
Patient characteristics	r	B	*P* value	B	*P* value	B	*P* value
Age, years	-0.155	-0.256	**0.006**	-0.222	**0.017**	-0.213	**0.020**
Male gender	-0.033	-1.528	0.554	-	-	-	-
BMI, kg x m^{-2}	0.074	0.083	0.185	-	-	-	-
Diabetes	-0.174	-9.245	**0.002**	-3.943	0.211	-3.105	0.331
NAFLD	-0.216	-9.951	**<0.001**	-4.014	0.196	-2.360	0.479
Liver stiffness, kPa	0.120	0.153	**0.032**	0.024	0.776	0.043	0.605
Histological fibrosis stage	0.028	0.461	0.621	-	-	-	-
Hepatic steatosis, %	-0.219	-0.171	**<0.001**	-0.124	**0.010**	-	-
CAP value, dB/m	-0.258	-0.079	**<0.001**	-	-	-0.057	**0.006**
Bilirubin, mg/L	0.124	0.756	**0.027**	0.393	0.303	0.378	0.320
ALT, U/L	0.003	0.001	0.957	-	-	-	-
AP, U/L	0.015	0.004	0.794	-	-	-	-
GGT, U/L	-0.010	-0.001	0.864	-	-	-	-

Multivariate (I+II): Due to multicollinearity, hepatic steatosis was included into model (I) and CAP value into model (II).

Abbreviations: CAP – controlled attenuation parameter; BMI – body mass index; NAFLD – non-alcoholic fatty liver disease; ALT – alanine aminotransferase; AP – alkaline phosphatase; γ-GT – gamma-glutamyltransferase

Supplementary Table 3. Comparison of different reliability criteria for **(A)** diagnosing hepatic steatosis (HS)≥S2 and **(B)** ≥S3.

A	**AUC of patients meeting this criterion**	**AUC of patients not meeting this criterion**	**Number of patients meeting this criterion**
CAP IQR <20	0.752 (0.607-0.896)	0.777 (0.718-0.835)	60 (18.8%)
CAP IQR <40	0.780 (0.715-0.844)	0.768 (0.677-0.860)	199 (62.4%)
CAP IQR <60	0.805 (0.752-0.857)	0.636 (0.470-0.802)	271 (85.0%)
CAP IQR <80	0.786 (0.735-0.837)	0.727 (0.381-1.000)	305 (95.6%)
CAP IQR/median <0.10	0.681 (0.576-0.786)	0.775 (0.702-0.847)	121 (37.9%)
CAP IQR/median <0.20	0.771 (0.712-.831)	0.686 (0.546-0.826)	239 (74.9%)
CAP IQR/median <0.30	0.777 (0.723-0.830)	0.741 (0.535-0.947)	288 (90.3%)
CAP IQR/median <0.40	0.779 (0.727-0.831)	0.643 (0.392-0.894)	304 (95.3%)
Stiffness IQR/median <0.3	0.782 (0.727-0.836)	0.821 (0.693-0.950)	272 (85.3%)

B	AUC of patients meeting this criterion	AUC of patients not meeting this criterion	Number of patients meeting this criterion
CAP IQR <20	0.616 (0.477-0.756)	0.800 (0.734-0.865)	60 (18.8%)
CAP IQR <40	0.718 (0.647-0.790)	0.819 (0.709-0.928)	199 (62.4%)
CAP IQR <60	0.755 (0.695-0.815)	0.760 (0.565-0.955)	271 (85.0%)
CAP IQR <80	0.758 (0.701-0.816)	-	305 (95.6%)
CAP IQR/median <0.10	0.613 (0.513-0.714)	0.829 (0.751-0.907)	121 (37.9%)
CAP IQR/median <0.20	0.716 (0.651-0.781)	0.724 (0.448-1.000)	239 (74.9%)
CAP IQR/median <0.30	0.738 (0.677-0.799)	-	288 (90.3%)
CAP IQR/median <0.40	0.753 (0.695-0.812)	-	304 (95.3%)
Stiffness IQR/median <0.3	0.760 (0.701-0.819)	0.765 (0.570-0.959)	272 (85.3%)

Abbreviations: CAP – controlled attenuation parameter; IQR – interquartile range; AUC – area under the receiver operating characteristic curve

Supplementary Table 4. Prevalence of hepatic steatosis according to CAP and liver histology.

	S0 (histology)	S1 (histology)	S2 (histology)	S3 (histology)	Sum
S0 (CAP)	75	35	12	2	124
S1 (CAP)	6	13	5	3	27
S2 (CAP)	5	9	9	5	28
S3 (CAP)	7	43	41	49	140
Sum	93	100	67	59	319

Abbreviations: CAP – controlled attenuation parameter

Supplementary Table 5. Diagnostic performance and proportion of discordant results between controlled attenuation parameter (CAP) and liver histology using the Youden's index derived cut-off (>246dB/m) for any hepatic steatosis (HS, i.e. ≥S1).

	AUC (95% CI)	**Prevalence of any steatosis (≥S1)**	**Discordance in presence of any steatosis (≥S1)**	***P* value**
Overall cohort	0.843 (0.798-0.887)	226 (70.8%)	67 (21.0%)	-
Cholestatic liver disease (PBC/PSC, n=18) vs. others	0.667 (0.376-0.957)	3 (17.6%)	4 (23.5%) vs. 63 (20.9%)	0.793
Autoimmune hepatitis (n=15) vs. others	0.620 (0.316-0.924)	5 (33.3%)	6 (40.0%) vs. 61 (20.1%)	0.064
BMI >30kg/m² (n=133) vs. ≤30kg/m²	0.871 (0.785-0.958)	118 (88.7%)	15 (11.3%) vs. 52 (28.0%)	**<0.001**
BMI >40kg/m² (n=40) vs. ≤40kg/m²	0.951 (0.877-1.000)	36 (90.0%)	3 (7.5%) vs. 64 (22.9%)	**0.025**
Diabetes (n=76) vs. others	0.925 (0.844-1.000)	70 (92.1%)	7 (9.2%) vs. 60 (24.7%)	**0.005**
F4 (n=79) vs. F0-3	0.831 (0.730-0.932)	55 (69.6%)	24 (30.4%) vs. 43 (17.9%)	**0.018**
≥10kPa liver stiffness (n=178) vs. <10kPa	0.850 (0.792-0.907)	118 (66.3%)	39 (58.2%) vs. 139 (55.2%)	0.655
≥15kPa liver stiffness (n=122) vs. <15kPa	0.820 (0.799-0.910)	76 (62.3%)	32 (26.2%) vs. 35 (17.8%)	0.071
≥20kPa liver stiffness (n=98) vs. <20kPa	0.780 (0.687-0.873)	59 (60.2%)	29 (29.6%) vs. 38 (17.2%)	**0.012**
ALT >2x sex-specific ULN (n=55) vs. <2xULN	0.867 (0.759-0.975)	37 (67.3%)	12 (21.8%) vs. 55 (20.8%)	0.870
γ-GT >2x sex-specific ULN (n=125) vs. <2xULN	0.823 (0.748-0.899)	79 (63.2%)	29 (23.2%) vs. 38 (19.6%)	0.439
CAP IQR <40dB/m (n=199) vs. ≥40dB/m	0.866 (0.812-0.920)	149 (74.9%)	35 (17.6%) vs. 32 (26.7%)	0.054
CAP IQR/median <0.3 (n=288) vs. ≥0.3	0.856 (0.792-0.886)	213 (74.0%)	54 (18.8%) vs. 13 (41.3%)	**0.003**

Abbreviations: CAP – controlled attenuation parameter; AUC – area under the receiver operating characteristic curve; 95%CI – 95% confidence interval; IQR – interquartile range; PBC – primary biliary cirrhosis; PSC – primary sclerosing cholangitis; BMI – body mass index; ALT – alanine aminotransferase; γ-GT – gamma-glutamyltransferase; IQR – interquartile range

Supplementary Table 6. Comparison of different reliability criteria for diagnosing any hepatic steatosis (≥S1) in **(A)** obese patients (BMI>30kg/m², n=133) and **(B)** patients with diabetes.

A	**AUC of patients meeting this criterion**	**AUC of patients not meeting this criterion**	**Number of patients meeting this criterion**
CAP IQR <20	-	0.848 (0.749-0.946)	27 (20.3%)
CAP IQR <40	0.865 (0.716-1.000)	0.847 (0.733-0.962)	79 (59.4%)
CAP IQR <60	0.882 (771-0.993)	0.788 (0.584-0.991)	111 (83.5%)
CAP IQR <80	0.857 (0.755-0.959)	0.778 (0.335-1.000)	127 (95.5%)
CAP IQR/median <0.10	0.883 (0.729-1.000)	0.826 (0.709-0.943)	62 (46.6%)
CAP IQR/median <0.20	0.862 (0.717-1.000)	0.680 (0.466-0.893)	109 (82.0%)
CAP IQR/median <0.30	0.879 (0.780-0.979)	0.533 (0.101-0.965)	125 (94.0%)
CAP IQR/median <0.40	0.868 (0.773-0.964)	0.250 (0.000-0.798)	129 (97.0%)
Stiffness IQR/median <0.3	0.881 (0.785-0.976)	0.804 (0.543-1.000)	113 (85.0%)

B	AUC of patients meeting this criterion	AUC of patients not meeting this criterion	Number of patients meeting this criterion
CAP IQR <20	-	0.917 (0.818-1.000)	23 (30.3%)
CAP IQR <40	0.930 (0.835-1.000)	0.889 (0.695-1.000)	56 (73.7%)
CAP IQR <60	0.926 (0.838-1.000)	-	71 (93.4%)
CAP IQR <80	0.934 (0.856-1.000)	-	75 (98.7%)
CAP IQR/median <0.10	-	0.870 (0.724-1.000)	43 (56.6%)
CAP IQR/median <0.20	0.909 (0.788-1.000)	0.750 (0.310-1.000)	69 (90.8%)
CAP IQR/median <0.30	0.923 (0.834-1.000)	1.000 (1.000-1.000)	74 (97.4%)
CAP IQR/median <0.40	0.934 (0.856-1.000)	-	75 (98.7%)
Stiffness IQR/median <0.3	0.923 (0.836-1.000)	1.000 (1.000-1.000)	69 (90.8%)

Abbreviations: CAP – controlled attenuation parameter; IQR – interquartile range; AUC – area under the receiver operating characteristic curve

Supplementary Table 7. Comparison of the diagnostic performance of CAP for diagnosing any hepatic steatosis(≥S1) in patients with non-alcoholic fatty liver disease (NAFLD) and viral hepatitis, stratified according to histological inflammation grades, applying NAFLD activity score (NAS) for NAFLD and METAVIR score for viral hepatitis.

		AUC (95%CI)	Number of patients
NAFLD	**No inflammation (0 points)**	0.665 (0.441-0.889)	36
	Mild inflammation (1 point)	0.890 (0.829-0.951)	115
	Moderate inflammation (2 points)	-	23
	Severe inflammation (3 points)	-	3
	Any inflammation (1-3 points)	0.896 (0.843-0.949)	141
Viral hepatitis	**No inflammation (0 points)**	0.639 (0.273-1.000)	13
	Mild inflammation (1 point)	0.747 (0.554-0.941)	27
	Moderate inflammation (2 points)	0.778 (0.447-1.000)	9
	Severe inflammation (3 points)	-	0

- indicates values that could not be computed; Abbreviations: CAP – controlled attenuation parameter; NAFLD – non-alcoholic fatty liver disease; NAS – NAFLD activity score; AUC – area under the receiver operating characteristic curve

Supplementary Table 8. Mean CAP values within the same stage of hepatic steatosis according to histology in patients with non-alcoholic fatty liver disease (NAFLD) and viral hepatitis, stratified according to histological inflammation grades, applying NAFLD activity score (NAS) for NAFLD and METAVIR score for viral hepatitis.

			Inflammation grade on histology				*P* value
			A0	**A1**	**A2**	**A3**	
NAFLD	**Steatosis stage on histology**	**S0 (n=11)**	246±73	250±27	-	-	0.768
		S1 (n=66)	283±72	302±74	316±77	-	0.552
		S2 (n=55)	300±44	322±59	323±55	-	0.728
		S3 (n=45)	256±6	336±36	341±35	356±44	**0.018**
Viral hepatitis		**S0 (n=27)**	201±49	212±47	216±57	-	0.841
		≥S1 (n=22)	203±70	262±63	257±62	-	0.286

- indicates values that could not be computed; Abbreviations: CAP – controlled attenuation parameter; NAFLD – non-alcoholic fatty liver disease; NAS – NAFLD activity score

Liver-FibroSTARD checklist (modified according to Bousier et al [2014])

Title/Abstract/ Keywords	**1. Identify the article as a study of diagnostic accuracy (recommend MeSH heading "sensitivity and specificity").**	
	1.1. Identify the article, especially in the title, as a study of the diagnostic performance of liver fibrosis/cirrhosis biomarker(s)/test(s).	
	1.2. Recommended key words (choose the most appropriate): "liver fibrosis", "cirrhosis", "diagnosis", "biomarker", "diagnostic test", "noninvasive diagnosis".	
Introduction	**2. State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups**.	
	In study aims, specify: 2.1. If the aim is to identify new marker(s)/develop new test(s), or to evaluate published marker(s)/test(s).	
	2.2. Whether the study is performed in a single or multiple cause(s) of chronic liver disease.	
	2.3. The reference used for fibrosis diagnosis in the study.	
	2.4. The diagnostic target used as the primary aim of the study and, if appropriate, other diagnostic targets used as secondary aims.	
Methods	Describe:	
Participants	**3. The study population: The inclusion and exclusion criteria*, setting, and locations* where data were collected.**	
	4. Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	
	4.1. State if healthy subjects without chronic liver disease are included or not in the study.	
	4.2. State if patients were selected by one abnormal or several discordant fibrosis test(s).	
	4.3. State if patients were selected according to the availability of reference or index test(s) result(s).	
	5. Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.	
	6. Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	
	6.1. The chronology between patient inclusion*, data collection (reference/index tests)*, and data analysis is well described.	
	6.2. Has the study population been previously used/published for the evaluation of the studied fibrosis test(s)?	
Test methods	**7. The reference standard and its rationale.**	

	8. Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	
	For the reference and index test(s), specify characteristics with sufficient detail to permit exact reoperation, when appropriate:	
	8.1. Center: standardization of procedures across centers.	
	8.2. Patient: fasting conditions*, time, posture, etc. (give information about the influence of conditions on the intra-individual variability).	
	8.3. Delay: time interval between reference and index test(s).	
	8.4. Material: technical specifications (name, generation, manufacturer, instrument), method of measurement, applicability (failure/reliability criteria)*. Specifically for liver biopsy, indicate material used per center, i.e., percutaneous/transjugular/other, needle diameter.	
	8.5. Biological samples: description of method of collection, transport, storage*.	
	8.6. Specify how the index tests were calculated	
	8.7. Specify how the risk for false negative/positive results was taken into account.	n.a.
	Specifically for liver biopsy:	
	8.8. How sample bias was limited: minimal biopsy size (length)*, number of portal tracts required, number of fragments.	
	8.9. Methods for histological assessment: human/automated reading*, local/central reading*, number and expertise of pathologists*, single/double reading*, consensus methods.	
	8.10. Scoring system used (Metavir, Ishak, Scheuer, etc.).	
	9. Definition of and rationale for the units, cut-offs*, and/or categories of the results of the index tests and the reference standard.	
	10. The number*, training and expertise* of the persons executing and reading the index tests and the reference standard.	
	11. Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	
Statistical methods	**12. State if the study is conducted on an intention-to-diagnose basis or if the analysis is per-protocol (i.e., with exclusion of failed/unreliable fibrosis test(s)/reference measurements).**	
	12.1. If intention-to-diagnose analysis, specify how failure and unreliable test(s)/reference are taken into account in the analysis.[a]	n.a.
	13. Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g., 95% confidence intervals).	
	Specify:	
	13.1. Detailed sample size calculation.	n.a.
	13.2. Statistical methods used to quantify uncertainty (e.g., 95% confidence intervals).	
	13.3. Control of multiple comparisons that increases type I error: multiple comparisons of tests (e.g. Bonferroni correction, etc.), multiple diagnostic targets.	
	13.4. Method for calculation of fibrosis test(s) diagnostic cut-offs.	
	13.5. Method for validation of new test(s) or new calculated diagnostic cut-off(s) (e.g., external validation set, internal validation by bootstrapping, etc.).	
	13.6. Method for control of center/operator effect.	n.a.

	13.7. Method for control of spectrum effect if unrepresentative prevalence of fibrosis stages (e.g., Obuchowski index, DANA, etc.).	n.a.
	13.8. Method for control of misclassification errors by the reference test.[b]	n.a.
	13.9. Use of a reference without gold standard.	n.a.
	13.10. Analysis of discordances between reference/index test(s).	
	14. Methods for calculating test reproducibility.	n.a.
Results	Report:	
Participants	**15. When study was performed, including beginning and end dates of recruitment.**	
	16. Clinical and demographic characteristics of the study population (e.g., age*, sex*, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centers).	
	16.1. For liver biopsy: size (length)*, number of portal tracts, number of fragments.	
	16.2. For index test(s): confounding factors that potentially influence the test(s) results (flare-up, inflammation, other liver lesions, intrinsic characteristics, etc.).	
	17. The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard*; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	
	17.1. If per-protocol analysis, report comparisons between patients excluded due to failed/unreliable test(s)/reference and patients with reliable fibrosis test(s)/reference.	n.a.
Test results	**18. Time-interval* between the index tests and the reference standard, and any treatment administered between.**	
	19. Distribution of severity of disease (define criteria) in those with the target condition*; other diagnoses in participants without the target condition.	
	19.1. Specify the prevalence* of the diagnostic condition (spectrum effect).	
	20. A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	
	20.1. Presentation of contingency tables, box/scatter plots.	
	21. Any adverse events from performing the index tests or the reference standard.	n.a.
Estimates	**22. Estimates of diagnostic accuracy* and measures of statistical uncertainty (e.g., 95% confidence intervals).**	
	22.1. Specify sensitivity* and specificity* with 95% confidence intervals; ROC analysis.	
	22.2. Analyzing discordances between fibrosis tests(s)/reference.[b]	
	23. How indeterminate results, missing data and outliers of the index tests were handled.	n.a.
	23.1. How missing/failure/unreliable results of index test(s)/reference were handled (intention-to-diagnose/per-protocol analysis).[a]	n.a.
	23.2. How outliers of the index tests were handled.	n.a.
	24. Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	
	25. Estimates of test reproducibility, if done.	n.a.
	26. Estimates of cost-benefit.	n.a.
Discussion	**27. Discuss the clinical applicability of the study findings.**	

	27.1. Discuss the representativeness of the study sample and recruiting centers (i.e., spectrum effect, etc.).	
	27.2. Discuss the interpretation of fibrosis test(s) results in clinical practice.	
	27.3. Discuss the clinical relevance of the study results.	

The Liver-FibroSTARD checklist summarizes the important information that must be present in the manuscripts of diagnostic studies on non-invasive tools for liver fibrosis evaluation. Compared to STARD, the Liver-

FibroSTARD checklist includes two additional items (#12 and #26) and 44 sub-items. The sub-items correspond to those proposals that clearly depicted, within the items, each of the particular features of diagnostic

studies on liver fibrosis tests. Finally, Liver-FibroSTARD presents as a complementary module of the STARD checklist.

Some items or sub/items include several criteria; ***major criteria are indicated by an asterisk (*).*** *Example: item #3: "The study population: The inclusion and exclusion criteria, setting, and locations where data were*

collected". If a major item is missing, the corresponding criterion has to be rated absent.

[a]*Items 12.1 and 23.1 are redundant but retained since they can be located in different paragraphs within an article.*

[b]*Items 13.10 and 22.2 are redundant but retained since they can be located in different paragraphs within an article.*

NOTE:
GREEN= addressed
YELLOW= partially addressed
n.a. = not applicable