Supplemental Methods

Study participant and design

Several individuals were prospectively recruited at the NAFLD Research Center including subject without NAFLD in the Twins and Family cohort and some normal controls based upon MRI-PDFF. Participants without NAFLD recruited in this cohort who underwent MRI-PDFF and CAP assessment within a six-month period were consecutively enrolled in the present study.

All participants underwent a standardized clinical research visit including detailed medical history, alcohol quantification using Skinner and Audit questionnaire, anthropometric exam, physical exam, and biochemical testing at the University of California at San Diego (UCSD) NAFLD Research Center (*9, 10, 45, 51-55*). After careful history, physical exam, and laboratory assessment all participants underwent advanced magnetic resonance imaging (MRI) based phenotyping and FibroScan assessment. NAFLD was assessed clinically and quantified by magnetic-resonance-imaging proton-density-fat-fraction (MRI-PDFF), which is an accurate, reproducible, and highly precise quantitative imaging based biomarker for liver fat assessment at UCSD MR3T Research Laboratory. Research visits and imaging procedures were performed the same day for each pair of twins, parent-offspring or siblings. This study was Health Insurance Portability and Accountability Act (HIPAA) compliant and was approved by the UCSD Institutional Review Board approval number: 111282. Informed written consent was obtained from each participant before enrolling in the study.

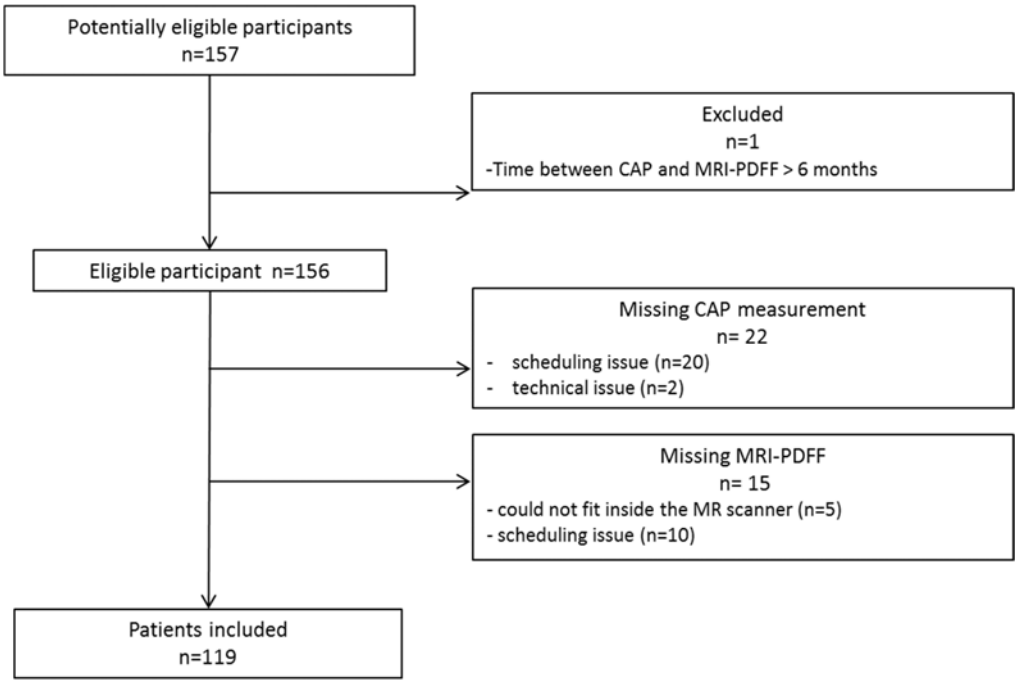
Inclusion and exclusion criteria

Patients were included if they were twins, siblings or parent-offspring at least 18 years old, willing and able to complete all research procedures and observations. For each twin pair, a detailed assessment of twinship status (ie, monozygotic (MZ) or dizygotic (DZ)) was obtained. The majority of twin-pairs (34) were diagnosed by their physician as either MZ or DZ by genetic testing. Furthermore, twin-ship status was confirmed by using a previously published questionnaire(*56*).

Participants were excluded from the study if they met any of the following criteria: significant alcohol intake (>10 g/day in females or >20 g/day in males) for at least 3 consecutive months over the previous 12 months or if the quantity of alcohol consumed could not be reliably ascertained; clinical or biochemical evidence of liver diseases other than NAFLD (eg, viral hepatitis, HIV, coeliac disease, cystic fibrosis, autoimmune hepatitis); metabolic and/or genetic liver disease (eg, Wilson's disease, haemochromatosis, polycystic liver disease, alpha-1-antitrypsin deficiency, dysbetalipoproteinaemia); clinical or laboratory evidence of systemic infection or any other clinical evidence of liver disease associated with hepatic steatosis; use of drugs known to cause hepatic steatosis (eg amiodarone, glucocorticoids, methotrexate, L-asparaginase and valproic acid) for at least 3 months in the last past 6 months; history of bariatric surgery; presence of systemic infectious illnesses; females who were pregnant or nursing at the time of the study; contraindications to MRI (eg metal implants, severe claustrophobia, body circumference greater than the imaging chamber); any other condition(s) which, based on the principal investigator's opinion, may significantly affect the participant's compliance, competence, or ability to complete the study.

Definition of NAFLD

NAFLD was defined as a MRI-PDFF $\geq$5% without any secondary causes of hepatic steatosis (see exclusion criteria listed above for details)



Supplemental Figure 1. Derivation of study cohort

157 participants were potentially eligible, 1 patient did not have CAP and MRI-PDFF within a period of 6 months, 22 participants were excluded due to missing CAP measurements and 15 participants were excluded due to missing MRI measurements. A total of 119 participants were included.

Supplemental Table 1. STARD checklist

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	3
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	3
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	4-5
	4	Study objectives and hypotheses	5
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	5-6
Participants	**6**	Eligibility criteria	6-7
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	6-7
	8	Where and when potentially eligible participants were identified (setting, location and dates)	6
	9	Whether participants formed a consecutive, random or convenience series	6
Test methods	**10a**	Index test, in sufficient detail to allow replication	8
	10b	Reference standard, in sufficient detail to allow replication	8-9
	11	Rationale for choosing the reference standard (if alternatives exist)	8
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	7
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	7
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	8
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	8
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	9
	15	How indeterminate index test or reference standard results were handled	9
	16	How missing data on the index test and reference standard were handled	9
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	9
	18	Intended sample size and how it was determined	9
RESULTS			
Participants	**19**	Flow of participants, using a diagram	10 &Supplemental Figure 1
	20	Baseline demographic and clinical characteristics of participants	10 & Table 1
	21a	Distribution of severity of disease in those with the target condition	10 & Table 1
	21b	Distribution of alternative diagnoses in those without the target condition	NA
	22	Time interval and any clinical interventions between index test and reference standard	8

Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Figure 1
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	10 - 11
	25	Any adverse events from performing the index test or the reference standard	NA
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	14
	27	Implications for practice, including the intended use and clinical role of the index test	14
OTHER INFORMATION			
	28	Registration number and name of registry	6
	29	Where the full study protocol can be accessed	Available upon request to the corresponding author
	30	Sources of funding and other support; role of funders	1

Supplemental Table 2. Diagnostic accuracy of CAP for the detection of hepatic steatosis as defined by MRI-PDFF≥15% and MRI-PDFF≥20%.

	AUROC (95%CI)	Cutoff (dB/m)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Secondary analysis: MRI-PDFF ≥ 15%						
CAP (dB/m)	0.77 (0.68-0.85)	313	65.6	70.1	44.7	84.7
Secondary analysis: MRI-PDFF ≥ 20%						
CAP (dB/m)	0.76 (0.66-0.86)	321	69.2	71.7	23.1	95.0

Supplemental Table 3. CAP value using XL compared to M probe stratified by hepatic fat content assessed by MRI-PDFF.

	MRI-PDFF <5%	**MRI-PDFF 5-10%**	**MRI-PDFF ≥10%**
M-probe CAP (mean +/-SD)	230.25 (57.7)	254.47 (46.60)	324.89 (34.70)
XL-probe CAP (mean +/- SD)	300.14 (35.14)	295.00 (47.00)	339.21 (28.24)
Statistical differences, p-value*	0.005	0.034	0.127

MRI-PDFF: magnetic resonance imaging proton-density fat fraction, CAP: controlled attenuation parameter, p-value determined by comparing the mean value of M and XL probe using an independent two-tailed t- test.