Title

Repeatability of MRI Biomarkers in Nonalcoholic Fatty Liver Disease: The NIMBLE Consortium.

Permalink

Journal

ISSN

Authors

Fowler, Kathryn J
Venkatesh, Sudhakar K
Obuchowski, Nancy
et al.

Publication Date

DOI

Peer reviewed

Repeatability of MRI Biomarkers in Nonalcoholic Fatty Liver Disease: The NIMBLE Consortium

Kathryn J. Fowler, MD* • *Sudhakar K. Venkatesh, MD** • *Nancy Obuchowski, PhD* • *Michael S. Middleton, MD, PhD* • *Jun Chen, PhD* • *Kay Pepin, PhD* • *Jessica Magnuson, BS* • *Kathy J. Brown, BS* • *Danielle Batakis, BS* • *Walter C. Henderson, BS* • *Sudha S. Shankar, PhD* • *Tania N. Kamphaus, PhD* • *Alex Pasek, MD* • *Roberto A. Calle, MD* • *Arun J. Sanyal, MD* • *Rohit Loomba, MD* • *Richard Ehman, MD* • *Anthony E. Samir, MD, PhD* • *Claude B. Sirlin, MD*** • *Sarah P. Sherlock, PhD***

Background: There is a need for reliable noninvasive methods for diagnosing and monitoring nonalcoholic fatty liver disease (NAFLD). Thus, the multidisciplinary Non-invasive Biomarkers of Metabolic Liver disease (NIMBLE) consortium was formed to identify and advance the regulatory qualification of NAFLD imaging biomarkers.

Purpose: To determine the different-day same-scanner repeatability coefficient of liver MRI biomarkers in patients with NAFLD at risk for steatohepatitis.

Materials and Methods: NIMBLE 1.2 is a prospective, observational, single-center short-term cross-sectional study (October 2021 to June 2022) in adults with NAFLD across a spectrum of low, intermediate, and high likelihood of advanced fibrosis as determined according to the fibrosis based on four factors (FIB-4) index. Participants underwent up to seven MRI examinations across two visits less than or equal to 7 days apart. Standardized imaging protocols were implemented with six MRI scanners from three vendors at both 1.5 T and 3 T, with central analysis of the data performed by an independent reading center (University of California, San Diego). Trained analysts, who were blinded to clinical data, measured the MRI proton density fat fraction (PDFF), liver stiffness at MR elastography (MRE), and visceral adipose tissue (VAT) for each participant. Point estimates and CIs were calculated using χ^2 distribution and statistical modeling for pooled repeatability measures.

Results: A total of 17 participants (mean age, 58 years ± 8.5 [SD]; 10 female) were included, of which seven (41.2%), six (35.3%), and four (23.5%) participants had a low, intermediate, or high likelihood of advanced fibrosis, respectively. The different-day same-scanner mean measurements were 13%–14% for PDFF, 6.6 L for VAT, and 3.15 kPa for two-dimensional MRE stiffness. The different-day same-scanner repeatability coefficients were 0.22 L (95% CI: 0.17, 0.29) for VAT, 0.75 kPa (95% CI: 0.6, 0.99) for MRE stiffness, 1.19% (95% CI: 0.96, 1.61) for MRI PDFF using magnitude reconstruction, 1.56% (95% CI: 1.26, 2.07) for MRI PDFF using complex reconstruction, and 19.7% (95% CI: 15.8, 26.2) for three-dimensional MRE shear modulus.

Conclusion: This preliminary study suggests that thresholds of 1.2%–1.6%, 0.22 L, and 0.75 kPa for MRI PDFF, VAT, and MRE, respectively, should be used to discern measurement error from real change in patients with NAFLD.

ClinicalTrials.gov registration no. NCT05081427

© RSNA, 2023

Supplemental material is available for this article.

Approximately 25% of the general population in many parts of the world have nonalcoholic fatty liver disease (NAFLD) (1) and 1.5%–6.5% of the population are affected by the more advanced form of the disease, nonalcoholic steatohepatitis (NASH) (2). Patients with NASH are at risk for liver-related and cardiovascular morbidity (3). The current reference standard for diagnosis of NAFLD/NASH is liver biopsy with histologic

assessment of hepatic fat, inflammation, ballooning, and fibrosis. However, longitudinal monitoring for disease using liver biopsy is not feasible due to the associated risk of complication and cost (4,5).

To overcome these limitations and to facilitate therapeutic trials, there is strong interest in noninvasive methods for identifying patients at risk for NASH. The Foundation for the National Institutes of Health, Non-invasive Biomarkers

Abbreviations

FIB-4 = fibrosis based on four factors, MRE = MR elastography, NAFLD = nonalcoholic fatty liver disease, NASH = nonalcoholic steatohepatitis, NIMBLE = Non-invasive Biomarkers of Metabolic Liver Disease, PDFF = proton density fat fraction, VAT = visceral adipose tissue

Summary

NIMBLE 1.2 was a prospective study that evaluated repeatability of MRI biomarkers of liver fat, fibrosis, and body composition in participants with nonalcoholic fatty liver disease to inform their longitudinal use.

Key Results

■ This prospective study included 17 participants at risk for nonalcoholic fatty liver disease who underwent multiple MRI examinations to determine the different-day same-scanner repeatability coefficients of potential imaging biomarkers for nonalcoholic steatohepatitis.

■ The repeatability coefficients were 0.22 L (95% CI: 0.17, 0.29) for visceral adipose tissue; 0.75 kPa (95% CI: 0.6, 0.99) and 19.7% (95% CI: 15.8, 26.2) for two- and three-dimensional MR elastography measures, respectively; and 1.19% (95% CI: 0.96, 1.61) and 1.56% (95% CI: 1.26, 2.07) for MRI proton density fat fraction using magnitude or complex reconstruction, respectively.

of Metabolic Liver Disease (NIMBLE) consortium is a multistakeholder public-private partnership that aims to standardize, compare, and validate circulating and imaging biomarkers for the noninvasive diagnosis and monitoring of NAFLD/NASH (6).

NIMBLE 1.2 was designed to assess the potential of using liver fat, liver stiffness, and body composition as imaging biomarkers for NASH. Modalities for measuring these biomarkers include MRI proton density fat fraction (PDFF) estimation, which occurs in two varieties (magnitude and complex reconstruction, where the former is a method that can be employed on any scanner and the latter requires commercial software and is not always available on all clinical scanners) and has been validated against pathology and spectroscopy as a highly accurate marker of liver fat (7–9). Additionally, MR elastography (MRE) can be used to estimate liver stiffness and has shown promise for diagnosing significant fibrosis (METAVIR score ≥F2) when compared with histopathology and other noninvasive methods (10–14). Due to the interplay between obesity, insulin resistance, and NAFLD in patients with metabolic syndrome (15), there has also been increasing interest in measuring body composition metrics, such as visceral adipose tissue (VAT).

While these noninvasive MRI markers show promise, further validation of their repeatability and reproducibility in patients with NAFLD is needed to qualify them as biomarkers for clinical trial use.

The primary objective of NIMBLE 1.2 was to evaluate the pooled different-day repeatability coefficient of MRI PDFF, MRE stiffness, and VAT at 1.5 T and 3 T. The study was designed around a central hypothesis that the upper 95% confidence bound of the pooled different-day repeatability coefficient of the imaging biomarkers would be less than 25%, on the basis of pre-existing knowledge and literature (16). A secondary aim of this study was to evaluate additional repeatability and reproducibility indexes for these imaging biomarkers and determine factors that may impact variability in measurements.
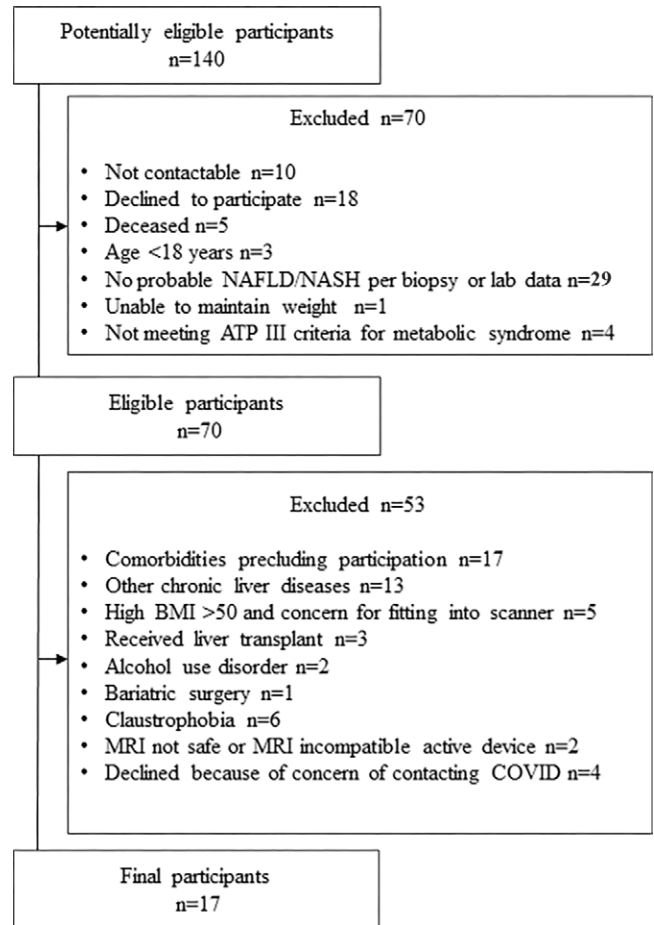

Figure 1: Flowchart shows study enrollment. ATP = adult treatment panel, BMI = body mass index, NAFLD = nonalcoholic fatty liver disease, NASH = nonalcoholic steatohepatitis.

Materials and Methods

This Health Insurance Portability and Accountability Act–compliant study was conducted following institutional review board approval of the protocol, and written informed consent was obtained from all participants.

Study Design and Participants

NIMBLE 1.2 is a single-center, prospective cross-sectional short-term study of adult participants who received a suspected or confirmed diagnosis of NAFLD at Mayo Clinic in Rochester, Minnesota from October 2021 to June 2022 (ClinicalTrials.gov registration no. NCT05081427). The study followed a protocol approved by the NIMBLE project team and in concordance with the Quantitative Imaging Biomarker Alliance (17).

Inclusion criteria were as follows: (a) age greater than or equal to 18 years; (b) probable NAFLD and NASH based on prior clinically performed biopsy within 36 months, or clinical or laboratory data within 3 months, consistent with NAFLD (alanine transaminase levels >30 U/L for male or >19 U/L for female participants, without other causes such as viral infection or alcohol); (c) able to have fibrosis status classified using the fibrosis based on four factors (FIB-4) index (18); (d) able
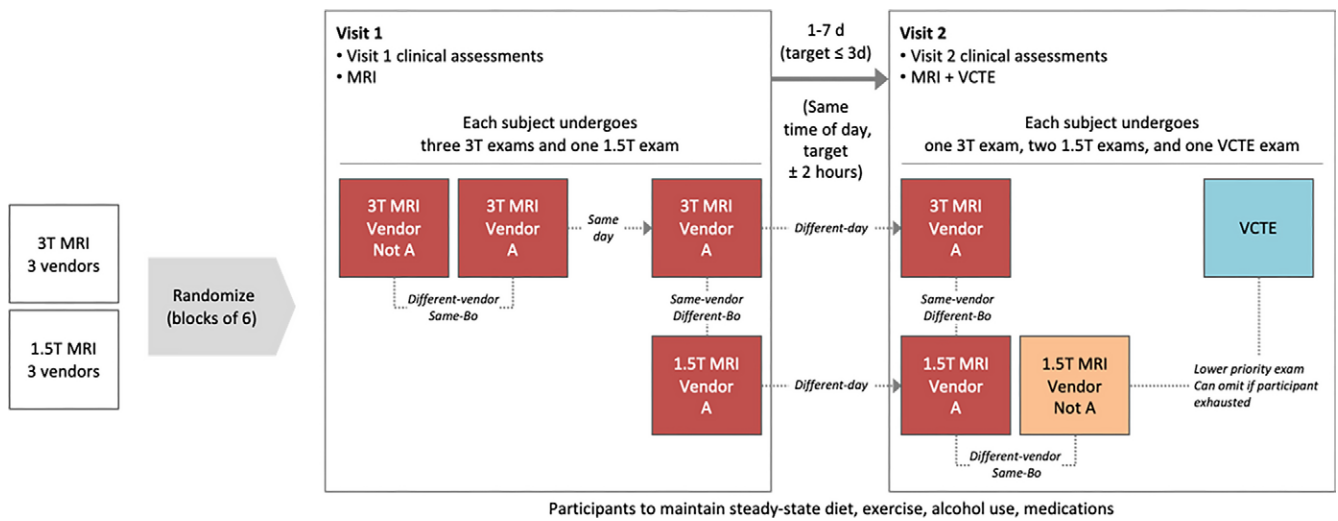
Figure 2: Schematic shows study procedures, including randomization to different blocks to ensure spread of participant data across all vendors and field strengths relatively equally. Each participant also underwent vibration-controlled transient elastography (VCTE) as part of the study. Imaging examinations were performed at approximately the same time each day and within 7 days of each other. Bo = field strength.
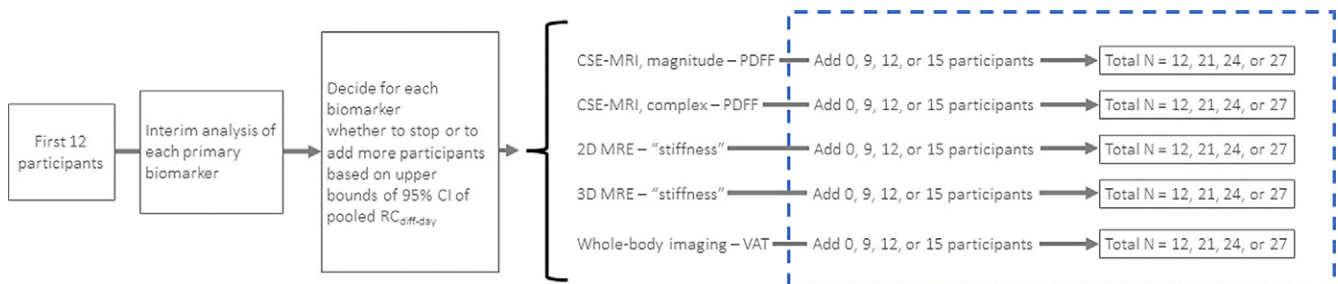

Figure 3: Schematic shows adaptive enrollment for determining the stopping rule for the interim analysis after enrollment of 12 participants, the results of which are used to determine if more participants need to be included in the final analysis. If the a priori upper bound 95% confidence threshold is met for each biomarker, no additional participants would be recruited. If a single biomarker or multiple biomarkers failed to meet the threshold, additional participants would be enrolled with another interim analysis planned upon enrollment of 21 and 24 participants. CSE = chemical shift encoding, MRE = MR elastography, PDFF = proton density fat fraction, $RC_{diff-day}$ = different-day repeatability coefficient, 3D = three-dimensional, 2D = two-dimensional, VAT = visceral adipose tissue.

and willing to participate, including maintaining diet, physical activity, alcohol consumption, and medication use unchanged from their usual regimen during the study period; and *(e)* meeting Adult Treatment Panel III criteria for metabolic syndrome (19). Exclusion criteria were as follows: *(a)* liver disease other than NAFLD; *(b)* excess alcohol consumption (>2 units per day for female or >3 units per day for male); *(c)* current diagnosis of drug-induced liver injury; *(d)* receiving a drug or placebo in a treatment trial during or within 30 days of the study period; *(e)* weight loss or gain of greater than 5 kg in the prior 3 months; and *(f)* other factors that might preclude study completion, including pregnancy, presence of implants such as pacemakers or defibrillators, or any other contraindication to MRI or vibration-controlled transient elastography (Fig 1).

The FIB-4 index, which considers patient age (years), aspartate aminotransferase levels (units per liter), alanine transaminase levels (units per liter), and platelet count (10^9 cells per liter), was used to ensure recruitment of participants across the spectrum of disease severity (17). Participants were categorized as follows: low likelihood of advanced fibrosis (FIB-4 index ≤1.3), intermediate likelihood of advanced fibrosis (FIB-4

index >1.3 to <2.67), or high likelihood of advanced fibrosis (FIB-4 index ≥2.67). In addition to imaging, enrolled participants completed questionnaires, anthropometric measurements, and blood collection. The biomarkers assessed were MRI PDFF, MRE stiffness, and whole-body VAT.

Imaging Procedures

MRI was performed with six different MRI scanners (HDxt 1.5 T and HDxt 3 T, GE HealthCare; Ingenia 1.5 T and Ingenia 3 T, Philips; and MAGNETOM Aera 1.5 T and Skyra 3 T, Siemens Healthineers), and vibration-controlled transient elastography was preformed using FibroScan (Echosens). Examinations were carried out as shown in Figure 2. Four MRI scans were acquired for each participant on day one, and two to three MRI scans were acquired for each participant on day two and performed around the same time of day within a 7-day interval. The acquired data were transmitted via AG Mednet Portal (AG Mednet) for central analysis, which was performed at the central reading center (University of California, San Diego) and by AMRA Medical. MRI protocol details are provided in Table S1. Participants were assigned to vendor and field strength according

Table 1: Characteristics of All Study Participants and Stratified according to Likelihood of Advanced Fibrosis

Characteristic	Low Likelihood of Advanced Fibrosis (n = 7)	Intermediate Likelihood of Advanced Fibrosis (n = 6)	High Likelihood of Advanced Fibrosis (n = 4)	All Participants (n = 17)
Age (y)	53.9 ± 10.2 (35–66)	62.2 ± 4.7 (56–69)	59.8 ± 7.9 (50–68)	58.2 ± 8.5 (35–69)
Sex*				
M	2 (28.6)	4 (66.7)	1 (25)	7 (41.2)
F	5 (71.4)	2 (33.3)	3 (75)	10 (58.8)
Waist circumference (inches)	41.6 ± 2.9 (38–47.5)	42.6 ± 2.0 (40.4–46.0)	44.6 ± 3.2 (42.0–49.0)	42.7 ± 2.8 (38–49)
Race*				
Asian	0	0	0	0
White	7 (100)	6 (100)	4 (100)	17 (100)
Other	0	0	0	0
Ethnicity*				
Hispanic or Latino	0	0	0	0
Not Hispanic or Latino	7 (100)	6 (100)	4 (100)	17 (100)
Smoking history*				
Never	5 (71.4)	3 (50)	3 (75)	11 (64.7)
Former	2 (28.6)	3 (50)	1 (25)	6 (35.3)
Current	0	0	0	0
Body mass index	32.5 ± 3.6 (27.5–39.1)	31.0 ± 2.4 (27.8–34.0)	35.8 ± 5.4 (31.4–43.6)	32.7 ± 4.0 (27.5–43.6)
FIB-4 index	0.9 ± 0.3 (0.49–1.24)	1.7 ± 0.2 (1.4–2.0)	4.5 ± 1.8 (2.86–7.0)	3.0 ± 2.5 (0.49–7.0)
Systolic BP (mm Hg)	123.3 ± 10.4 (105–136)	126.5 ± 8.4 (114–139)	117.5 ± 11.2 (104–129)	123.1 ± 9.9 (104–139)
Diastolic BP (mm Hg)	76.9 ± 8.0 (63–84)	78.3 ± 6.7 (67–83)	73.8 ± 3.8 (70–77)	76.6 ± 6.6 (63–84)
AST level (U/L)	28.4 ± 7.7 (15–38)	39.8 ± 18.1 (14–66)	88.8 ± 46.0 (53–153)	46.6 ± 33.6 (14–153)
ALT level (U/L)	41.4 ± 19.9 (21–68)	51.5 ± 28.0 (10–87)	83.0 ± 43.5 (40–131)	54.8 ± 32.1 (10–131)
Plasma glucose level (mg/dL)	89.7 ± 27.4 (35–126)	117.5 ± 28.6 (93–172)	140.3 ± 13.5 (126–155)	111.4 ± 31.6 (35–172)
Plasma urea nitrogen level (mg/dL)	14.7 ± 3.0 (10–18)	15.8 ± 3.5 (9–18)	16.0 ± 2.9 (13–19)	15.4 ± 3.0 (9–19)
Total cholesterol level (mg/dL)†	158.7 ± 20.6 (133–182)	160.3 ± 47.9 (101–237)	163.5 ± 64.2 (106–247)	160.5 ± 41.7 (101–247)
LDL level (mg/dL)	84.0 ± 22.1 (56–111)	105.0 ± 53.2 (46–180)	86.5 ± 53.8 (47–163)	92.0 ± 41.3 (46–180)
HDL level (mg/dL)	47.3 ± 10.4 (35–64)	39.0 ± 6.4 (31–49)	49.8 ± 19.2 (32–76)	44.9 ± 12.0 (31–76)
Triglyceride level (mg/dL)	148.6 ± 66.3 (82–261)	204.2 ± 59.7 (147–303)	135.5 ± 58.0 (93–218)	165.1 ± 65.6 (82–303)
Platelet count (10⁹/L)	289.4 ± 36.5 (224–330)	196.5 ± 25.7 (165–236)	137.5 ± 57.1 (70–208)	220.9 ± 73.0 (70–330)
Median SWS (m/sec)	1.77 ± 0.34 (1.21–2.23)	1.96 ± 0.29 (1.54–2.41)	2.54 ± 0.91 (1.59–3.61)	2.02 ± 0.57 (1.21–3.61)
CAP (dB/m)	297.9 ± 46.3 (245–386)	323.0 ± 28.5 (283–358)	291.5 ± 35.8 (261–330)	305.2 ± 38.6 (245–386)
AUDIT result*				
Abstinence	0	3 (50)	3 (75)	6 (35.3)
Low risk	7 (100)	3 (50)	1 (25)	11 (64.7)

Note.—Data are means ± SDs, with ranges in parentheses, for continuous data. Body mass index is calculated as weight in kilograms divided by height in meters squared. The likelihood of advanced fibrosis was determined using the FIB-4 index in which values less than or equal to 1.3 indicate low likelihood, values greater than 1.3 to less than 2.67 indicate intermediate likelihood, and values greater than or equal to 2.67 indicate high likelihood. AUDIT is a standardized survey tool (score 0–40) used to quantify alcohol consumption, whereby a score of 0 indicates no alcohol consumption concerns and a score 1–7 suggests low-risk consumption. ALT = alanine transaminase, AST = aspartate transferase, AUDIT = Alcohol Use Disorders Identification Test, BP = blood pressure, CAP = controlled attenuation parameter, FIB-4 = fibrosis based on four factors, HDL = high-density lipoprotein, LDL = low-density lipoprotein, SWS = shear-wave speed.

* Data are numbers of participants, with percentages in parentheses, for categorical data.

† One participant with a low likelihood of advanced fibrosis is missing data for this variable.

to a block randomization scheme. The site qualification process is described in Appendix S1.

Biomarker Analysis

MRI PDFF and MRE data were evaluated using Horos (Purview) Digital Imaging and Communications in Medicine image viewer software and MRE Quant (Resondant)

software. The three-dimensional MRE data underwent multimodal direct inversion reconstruction by Resondant prior to transmission to the central reading center for analysis (20). All trained image analysts (D.B., W.H.), with at least 1 year of experience, were blinded to participant clinical information, including the FIB-4 index. Analysis was performed in accordance with the NIMBLE 1.2 Imaging Review Charter
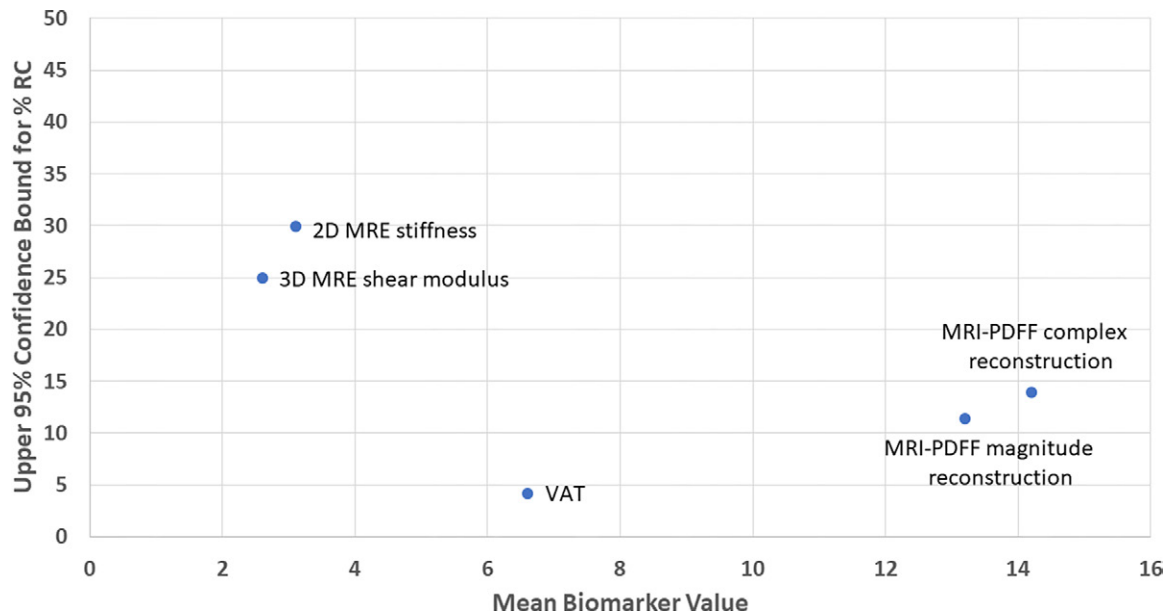
Figure 4: Point estimate graph of the upper 95% confidence bounds for different-day repeatability coefficient percentages (%RC) for all imaging biomarkers shows the relative spread between biomarkers. MRE = MR elastography, PDFF = proton density fat fraction, 3D = three-dimensional, 2D = two-dimensional, VAT = visceral adipose tissue.

by placing nine regions of interest in each liver segment for PDFF and a large region of interest on four axial sections while avoiding the outer 1 cm of liver and including regions of parallel wave propagation within the 95% CI maps for MRE, and according to good clinical practice. After each image set was analyzed (including mean liver stiffness measured in kilopascals and mean liver PDFF percentage), a second analyst performed a quality inspection over-read to confirm the results. VAT (liters) and other body composition metrics were analyzed by AMRA Medical (21), with anonymization of other biomarker and participant data.

Statistical Analysis

The statistical analysis plan was developed by the study statistician (N.O.) and approved by the NIMBLE project team prior to study initiation. The study was designed around a central hypothesis that the upper 95% confidence bound of the pooled different-day repeatability coefficient of the imaging biomarkers would be less than 25%, on the basis of pre-existing knowledge and literature (16). For each performance estimate, 95% upper bounds were calculated based on a χ^2 distribution and, to account for clustered data at the participant level when appropriate, by both a generalized linear model with generalized estimating equations and bootstrapping. A Monte Carlo simulation study was conducted to determine the sample size for paired test and retest observations to achieve an 80% power with 5% type 1 error rate assuming normally distributed data and, conservatively, assuming interscanner variability in the different-day repeatability coefficients of 14%, 18%, and 22% across scanners. An adaptive enrollment scheme was used in which an interim analysis was conducted after the enrollment of 12 participants, the results of which would be used to determine if more participants needed to be included in the final analysis (Fig 3).

The different-day repeatability coefficient for the imaging biomarkers of interest was calculated by pooling data for each field strength across all scanners, details of which can be found in Appendix S1. The stopping point for the interim analysis was an upper bound for different-day repeatability coefficient values less than or equal to 24.1% for each biomarker. For the final analysis, if the upper 95% confidence bound was less than 24.1%, then it was concluded that the biomarker met the prespecified precision criterion of greater than 25%. Similar methods were used for the additional repeatability and reproducibility metrics (22). Linear regression of within-participant SD was also performed for each biomarker and is reported in Appendix S1 and Figures S1 and S2. Finally, a multivariable linear regression model was constructed to assess the effects of FIB-4 index, sex, body mass index (calculated as weight in kilograms divided by height in meters squared), and waist circumference on precision metrics, with results reported in Appendix S1 and Figure S3. Generalized estimating equations were used to handle the clustered data.

$P < .05$ was considered indicative of a statistically significant difference. All descriptive and inferential statistics were performed using SAS version 9.4 (SAS Institute) by the statistician, who was not part of the data acquisition or curation process. Descriptive statistics were expressed as counts and frequencies or means and SDs.

Results

Participant Characteristics

A total of 17 participants (mean age, 58 years ± 8.5 [SD]; 10 female, seven male) were included. Participants had a mean body mass index of 32.7 (range, 27.5–43.6) and a mean FIB-4 index of 3 (range, 0.49–7.0) (Table 1). All participants completed all study procedures. Although the interim analysis (which included
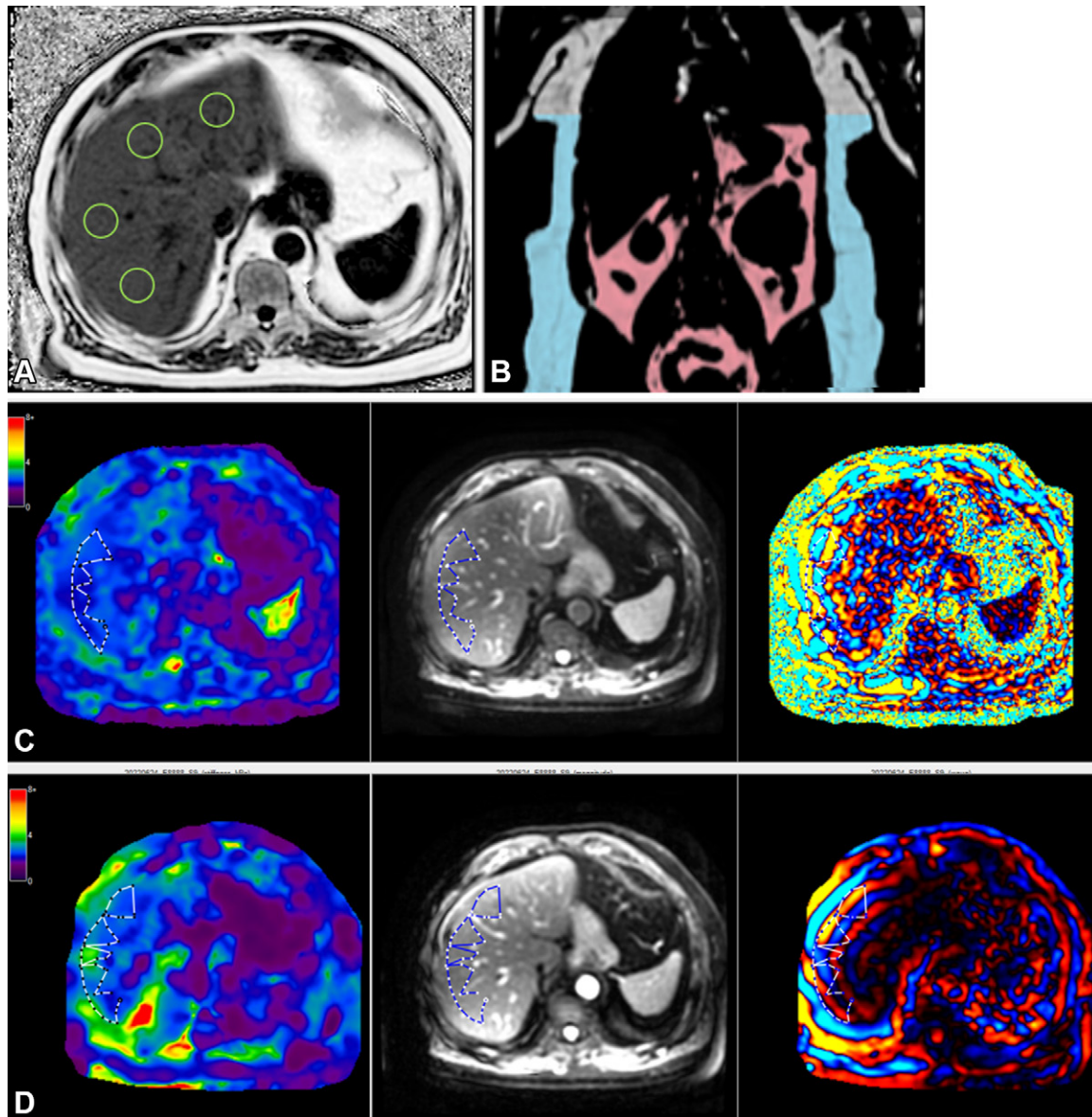
Figure 5: MRI in a 65-year-old male participant with low likelihood of advanced fibrosis according to the fibrosis based on four factors (FIB-4) index. **(A)** Axial MRI proton density fat fraction map shows regions of interest (green circles) drawn in each Couinaud segment while avoiding any vessels or lesions, as per the study protocol. **(B)** Coronal image shows body composition with autosegmented visceral (pink) and subcutaneous (blue) adipose tissue. **(C)** Left to right: Axial elastogram, magnitude, and wave images obtained at two-dimensional MR elastography (MRE) show placement of the region of interest (dotted outline) in the region of parallel wave propagation, avoiding the outer edge of the liver. **(D)** Left to right: Axial elastogram, magnitude, and wave images obtained at three-dimensional MRE similarly show the region of interest (dotted outline) in a region of parallel wave propagation, avoiding the outer margin of the liver. Color bars alongside elastograms indicate a range of kilopascals from low (purple) to high (red).

only the first 12 participants enrolled) met the stopping rule based on the primary end point analysis, an additional five participants were enrolled while awaiting interim analysis results. Thus, data presented herein include all 17 participants. During the study period, two participants were identified as having overt cirrhosis, defined as cross-sectional imaging evidence of cirrhosis and portal hypertension (23). To evaluate the potential impact of overt cirrhosis on repeatability, a separate analysis was performed from which these participants were excluded. There were no reported changes in medication, alcohol use, or exercise routines between the two visits for any of the study participants.

Primary End Point Analysis

The relative precision across all biomarkers is shown in Figure 4, and a sample placement of regions of interest is shown in Figure 5 for PDFF and MRE analyses. Figure 6 shows an example case of a high FIB-4 index at MRE. The results of the analysis without the two participants demonstrating overt cirrhosis were similar, with repeatability coefficients of 0.22 L, 0.65 kPa, and 1.3%–1.5% for VAT, MRE, and PDFF, respectively (Table S2). The repeatability coefficient percentages for hypothesis testing are provided in Table 2. Point estimates for pooled different-day repeatability coefficients for the imaging biomarkers were as follows: 1.19% (95% CI: 0.96, 1.61) for
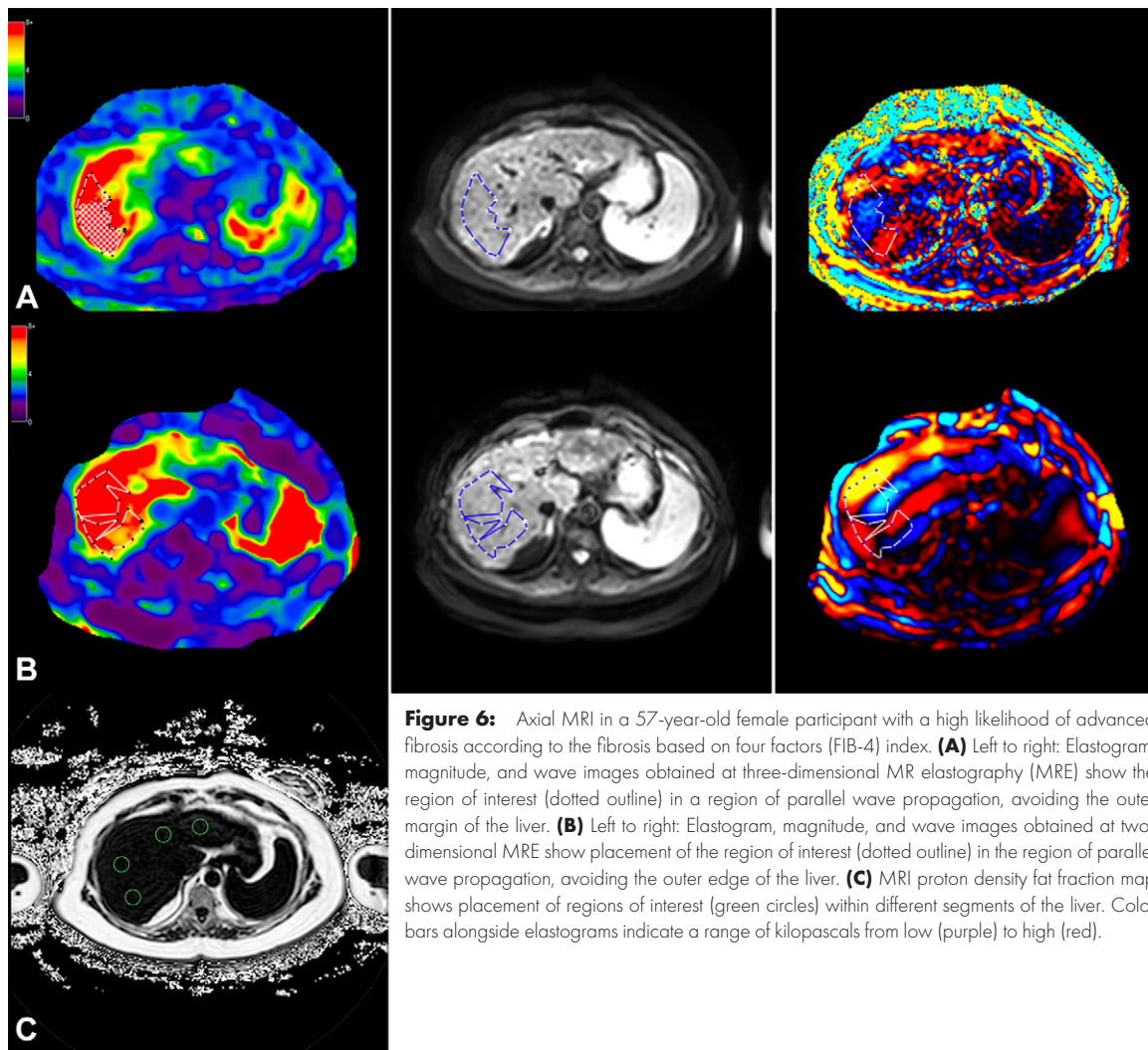
Figure 6: Axial MRI in a 57-year-old female participant with a high likelihood of advanced fibrosis according to the fibrosis based on four factors (FIB-4) index. **(A)** Left to right: Elastogram, magnitude, and wave images obtained at three-dimensional MR elastography (MRE) show the region of interest (dotted outline) in a region of parallel wave propagation, avoiding the outer margin of the liver. **(B)** Left to right: Elastogram, magnitude, and wave images obtained at two-dimensional MRE show placement of the region of interest (dotted outline) in the region of parallel wave propagation, avoiding the outer edge of the liver. **(C)** MRI proton density fat fraction map shows placement of regions of interest (green circles) within different segments of the liver. Color bars alongside elastograms indicate a range of kilopascals from low (purple) to high (red).

MRI PDFF using magnitude reconstruction, 1.56% (95% CI: 1.26, 2.07) for MRI PDFF using complex reconstruction, 0.75 kPa (95% CI: 0.60, 0.99) for liver stiffness measured with two-dimensional MRE, and 0.22 L (95% CI: 0.17, 0.29) for VAT (Table 2). For three-dimensional MRE, the repeatability coefficient percentage was 19.7% (95% CI: 15.8, 26.2).

Discussion

There is need for reliable noninvasive methods for diagnosing and monitoring nonalcoholic fatty liver disease (NAFLD). Thus, the multidisciplinary Non-invasive Biomarkers of Metabolic Liver Disease (NIMBLE) consortium was formed to identify and advance the regulatory qualification of NAFLD imaging biomarkers. The purpose of NIMBLE 1.2 was to determine the different-day same-scanner repeatability coefficient of liver MRI biomarkers in participants with NAFLD at risk for steato-hepatitis. The point estimates for repeatability coefficients were 1.19% (95% CI: 0.96, 1.61) for MRI proton density fat fraction (PDFF) using magnitude reconstruction, 1.56% (95% CI: 1.26, 2.07) for MRI PDFF using complex reconstruction, 0.75 kPa (95% CI: 0.60, 0.99) for two-dimensional MR elastography (MRE) stiffness, and 0.22 L (95% CI: 0.17, 0.29) for visceral

adipose tissue. The repeatability coefficient percentage for three-dimensional MRE was 19.7% (95% CI: 15.8, 26.2).

The present study extends the existing knowledge of repeatability and reproducibility for MRI biomarkers. Prior studies have evaluated the test-retest repeatability of MRI PDFF across vendors and field strengths in phantoms (24), using different technique parameters (25), and across different times of day to capture diurnal variations (26). Our results agree with prior reports that relative measurement differences up to 10% may occur due to measurement variance, with repeatability coefficients of 1.2%–1.6% in our study. Similarly, prior studies evaluating the test-retest repeatability of liver stiffness measurements have shown that up to 37% of true changes may be masked by measurement variance in patients with liver disease (27,28), with a smaller study in healthy volunteers and patients with hepatitis C demonstrating within-subject coefficients of variation ranging 6%–11% (29). Of note, the MRE studies referenced were performed in patients with other etiologies of liver disease, primarily hepatitis C, and not NAFLD/NASH. Alternatively, our study provides rigorous data on the reproducibility of quantitative MRI noninvasive tests in the relevant population of participants with NAFLD/NASH across different days, scanners, and

Table 2: Primary and Secondary End Point Analyses for Imaging Biomarkers in All Participants

Analysis and Biomarker	Mean Measurement	Within-Participant SD	Repeatability Coefficient	Upper 95% CI (%) (Primary Hypothesis)
Primary analysis				
Different-day same-scanner precision				
VAT volume (liters)	6.60	0.08	0.22 (0.17, 0.29)	4.2
2D MRE stiffness (kPa)	3.15	0.27	0.75 (0.60, 0.99)	29.9
Magnitude PDFF (%)	13.19	0.43	1.19 (0.96, 1.61)	11.4
Complex PDFF (%)	14.16	0.56	1.56 (1.26, 2.07)	13.9
3D MRE shear modulus (kPa)	2.63	0.07*	19.7% (15.8, 26.2)[†]	25
Secondary analysis				
Same-day, same field-strength, same-scanner precision				
VAT volume (liters)	6.55	0.06	0.16 (0.12, 0.24)	NA
2D MRE stiffness (kPa)	3.16	0.19	0.53 (0.40, 0.80)	NA
Magnitude PDFF (%)	12.27	0.16	0.44 (0.33, 0.66)	NA
Complex PDFF (%)	13.4	0.52	1.44 (1.08, 2.16)	NA
3D MRE shear modulus (kPa)	2.68	0.05*	12.4% (9.2, 18.9)[†]	NA
Same-day, same-vendor, different field-strength precision				
VAT volume (liters)	6.71	0.11	0.30 (0.22, 0.47)[‡]	NA
2D MRE stiffness (kPa)	2.99	0.19	0.53 (0.39, 0.83)[‡]	NA
Magnitude PDFF (%)	13.81	0.80	2.21 (1.63, 3.41)[‡]	NA
Complex PDFF (%)	14.73	0.98	2.71 (2.00, 4.19)[‡]	NA
3D MRE shear modulus (kPa)	2.48	0.06*	16.0% (11.7, 25.2)[†]	NA
Same-day, different-vendor, same field-strength precision				
VAT volume (liters)	6.56	0.15	0.41 (0.31, 0.61)[‡]	NA
2D MRE stiffness (kPa)	3.16	0.21	0.59 (0.45, 0.89)[‡]	NA
Magnitude PDFF (%)	13.61	0.71	1.97 (1.48, 2.95)[‡]	NA
Complex PDFF (%)	14.59	1.27	3.51 (2.63, 5.26)[‡]	NA
3D MRE shear modulus (kPa)	2.67	0.09*	26.7% (20.0, 40.1)[†]	NA

Note.—Data in parentheses are 95% CIs. Primary hypothesis testing was for estimating the 95% CI upper bound. The interim analysis was performed after the first 12 participants with values for each primary biomarker met the predefined threshold of less than 25%; however, the final cohort comprised a total of 17 participants. MRE = MR elastography, NA = not applicable, PDFF = proton density fat fraction, 3D = three-dimensional, 2D = two-dimensional, VAT = visceral adipose tissue.

* Data are within-participant coefficients of variation rather than within-participant SDs.

[†] For 3D MRE shear modulus, the repeatability coefficient percentage or reproducibility coefficient percentage is reported rather than an absolute value.

[‡] Data are reproducibility coefficients rather than repeatability coefficients.

field strengths. While PDFF has already been proposed as an end point for NASH clinical trials that assess changes in liver fat as a primary measure of drug efficacy (30), our results suggest that changes of 1.2%–1.6% are the thresholds that may be considered for estimating true change and power sample size for future studies.

Few studies have focused on the reproducibility or repeatability of body composition metrics. A study of 18 healthy volunteers assessed using a single MRI scanner reported a repeatability coefficient for VAT of 0.13 L (31). In another prospective cross-sectional study of 20 participants who were assessed using automated MRI from a single vendor, VAT measurements showed intra- and interexamination coefficients of variation of 3.3% and 3.6%, respectively (32). We found that VAT measurements are reproducible across vendors, field strengths, and temporal assessments, with a repeatability coefficient of 0.22 L, which will further support large-scale clinical trial use (33).

In addition, we evaluated the potential impact of participant-level factors of sex, body mass index, and severity of liver fibrosis (estimated according to the FIB-4 index) on measurement variation and found associations of BMI with repeatability coefficients for VAT, PDFF, and MRE. The impact of obesity on diagnostic accuracy of imaging biomarkers in NAFLD/NASH has been studied previously, with comparative studies showing that MRI may be less impacted by obesity than other imaging modalities, such as vibration-controlled transient elastography (34). Another study showed that as body mass index increased, discordance in fibrosis stage increased between MRE and vibration-controlled transient elastography (35). While it appears that body mass index may impact the performance of some

noninvasive tests, our study was not powered to examine this relationship or the underlying biologic mechanisms.

Our study had several limitations. First, we recruited participants with NAFLD who were at risk for NASH but had not been diagnosed with NASH. These participants have relatively low disease severity and so our results should be applied with caution in different patient populations, such as those with known advanced disease. Second, while we assessed potential predictors of variability, our study was not powered for this analysis. Finally, our sample was not diverse with regards to race and ethnicity and may be prone to selection bias. Future studies that include more individuals are needed to examine potentially modifiable predictors of variability, such as the alteration of specific sequence parameters, different operators, diurnal variation, or fasting state. Finally, other MRI noninvasive tests that have shown potential value in assessing NAFLD/NASH, such as R2*, additional three-dimensional MRE markers, and additional body composition metrics, were not analyzed in the present study. These data have been collected and will be reported in a future study.

In conclusion, our study assessing potential MRI biomarkers in participants with nonalcoholic fatty liver disease/nonalcoholic steatohepatitis showed that a difference of at least 0.22 L for visceral adipose tissue volume, 1.2%–1.6% for MRI proton density fat fraction, and 0.75 kPa for MR elastography may constitute real change not attributable to measurement error. Future studies to establish clinically relevant cutoff values for clinical trial enrichment and to further validate the biomarkers in combination with circulating markers are needed to advance the field, and are planned as next stages in the NIMBLE project plan.

Author contributions: Guarantors of integrity of entire study, **K.J.F., M.S.M., D.B., W.C.H., S.P.S.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **K.J.F., S.K.V., M.S.M., J.C., D.B., W.C.H., T.N.K., A.J.S., R.E., C.B.S.**; clinical studies, **K.J.F., M.S.M., J.C., K.P., J.M., K.J.B., D.B., W.C.H., S.S.S., R.L., R.E., A.E.S., C.B.S., S.P.S.**; experimental studies, **K.J.F., J.C., D.B., W.C.H., S.S.S., R.E., S.P.S.**; statistical analysis, **K.J.F., N.O., J.C., D.B., W.C.H.**; and manuscript editing, **K.J.F., S.K.V., N.O., M.S.M., J.C., K.P., D.B., W.C.H., S.S.S., T.N.K., A.P., R.A.C., A.J.S., R.L., R.E., A.E.S., C.B.S., S.P.S.**

References

1. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 2016;64(1):73–84.

2. Bondini S, Kleiner DE, Goodman ZD, Gramlich T, Younossi ZM. Pathologic assessment of non-alcoholic fatty liver disease. Clin Liver Dis 2007;11(1):17–23, vii.

3. Younossi Z, Anstee QM, Marietti M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol 2018;15(1):11–20.

4. Khalifa A, Rockey DC. The utility of liver biopsy in 2020. Curr Opin Gastroenterol 2020;36(3):184–191.

5. Seeff LB, Everson GT, Morgan TR, et al. Complication rate of percutaneous liver biopsies among persons with advanced chronic liver disease in the HALT-C trial. Clin Gastroenterol Hepatol 2010;8(10):877–883.

6. Sanyal AJ, Shankar SS, Calle RA, et al. Non-invasive biomarkers of nonalcoholic steatohepatitis: the FNIH NIMBLE project. Nat Med 2022;28(3):430–432.

7. Heba ER, Desai A, Zand KA, et al. Accuracy and the effect of possible subject-based confounders of magnitude-based MRI for estimating hepatic proton density fat fraction in adults, using MR spectroscopy as reference. J Magn Reson Imaging 2016;43(2):398–406.

8. Noureddin M, Lam J, Peterson MR, et al. Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials. Hepatology 2013;58(6):1930–1940.

9. Kim HJ, Cho HJ, Kim B, et al. Accuracy and precision of proton density fat fraction measurement across field strengths and scan intervals: A phantom and human study. J Magn Reson Imaging 2019;50(1):305–314.

10. Cui J, Heba E, Hernandez C, et al. Magnetic resonance elastography is superior to acoustic radiation force impulse for the Diagnosis of fibrosis in patients with biopsy-proven nonalcoholic fatty liver disease: A prospective study. Hepatology 2016;63(2):453–461.

11. Cui J, Ang B, Haufe W, et al. Comparative diagnostic accuracy of magnetic resonance elastography vs. eight clinical prediction rules for non-invasive diagnosis of advanced fibrosis in biopsy-proven non-alcoholic fatty liver disease: a prospective study. Aliment Pharmacol Ther 2015;41(12):1271–1280.

12. Taouli B, Serfaty L. Magnetic resonance imaging/elastography is superior to transient elastography for detection of liver fibrosis and fat in nonalcoholic fatty liver disease. Gastroenterology 2016;150(3):553–556.

13. Kim D, Kim WR, Talwalkar JA, Kim HJ, Ehman RL. Advanced fibrosis in nonalcoholic fatty liver disease: noninvasive assessment with MR elastography. Radiology 2013;268(2):411–419.

14. Loomba R, Wolfson T, Ang B, et al. Magnetic resonance elastography predicts advanced fibrosis in patients with nonalcoholic fatty liver disease: a prospective study. Hepatology 2014;60(6):1920–1928.

15. Tejani S, McCoy C, Ayers CR, et al. Cardiometabolic health outcomes associated with discordant visceral and liver fat phenotypes: insights from the Dallas Heart Study and UK Biobank. Mayo Clin Proc 2022;97(2):225–237.

16. Serai SD, Obuchowski NA, Venkatesh SK, et al. Repeatability of MR Elastography of Liver: A Meta-Analysis. Radiology 2017;285(1):92–100.

17. QIBA MR Biomarker Committee. MR Elastography of the Liver, Quantitative Imaging Biomarkers Alliance. Profile Stage: Technically Confirmed. https://qibawiki.rsna.org/images/5/54/MRE-QIBAProfile-2022-02-14-TECHNICALLY-CONFIRMED.pdf. Published February 14, 2022. Accessed February 15, 2023.

18. Sterling RK, Lissen E, Clumeck N, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology 2006;43(6):1317–1325.

19. Grundy SM, Cleeman JI, Daniels SR, et al. Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. Circulation 2005;112(17):2735–2752.

20. Silva AM, Grimm RC, Glaser KJ, et al. Magnetic resonance elastography: evaluation of new inversion algorithm and quantitative analysis method. Abdom Imaging 2015;40(4):810–817.

21. Borga M, Thomas EL, Romu T, et al. Validation of a fast method for quantification of intra-abdominal and subcutaneous adipose tissue for large-scale human studies. NMR Biomed 2015;28(12):1747–1753.

22. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res 2015;24(1):27–67.

23. Siddiqui MS, Harrison SA, Abdelmalek MF, et al. Case definitions for inclusion and analysis of endpoints in clinical trials for nonalcoholic steatohepatitis through the lens of regulatory science. Hepatology 2018;67(5):2001–2012.

24. Zhao R, Hernando D, Harris DT, et al. Multisite multivendor validation of a quantitative MRI and CT compatible fat phantom. Med Phys 2021;48(8):4375–4386.

25. Johnson BL, Schroeder ME, Wolfson T, et al. Effect of flip angle on the accuracy and repeatability of hepatic proton density fat fraction estimation by complex data-based, T1-independent, T2*-corrected, spectrum-modeled MRI. J Magn Reson Imaging 2014;39(2):440–447.

26. Colgan TJ, Van Pay AJ, Sharma SD, Mao L, Reeder SB. Diurnal variation of proton density fat fraction in the liver using quantitative chemical shift encoded MRI. J Magn Reson Imaging 2020;51(2):407–414.

27. Hines CDG, Bley TA, Lindstrom MJ, Reeder SB. Repeatability of magnetic resonance elastography for quantification of hepatic stiffness. J Magn Reson Imaging 2010;31(3):725–731.

28. Lee Y, Lee JM, Lee JE, et al. MR elastography for noninvasive assessment of hepatic fibrosis: reproducibility of the examination and reproducibility and repeat-ability of the liver stiffness value measurement. J Magn Reson Imaging 2014;39(2):326–331.

29. Shire NJ, Yin M, Chen J, et al. Test-retest repeatability of MR elastography for noninvasive liver fibrosis assessment in hepatitis C. J Magn Reson Imaging 2011;34(4):947–955.

30. 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(2):763–772.

31. Borga M, Ahlgren A, Romu T, Widholm P, Dahlqvist Leinhard O, West J. Reproducibility and repeatability of MRI-based body composition analysis. Magn Reson Med 2020;84(6):3146–3156.

32. Middleton MS, Haufe W, Hooker J, et al. Quantifying Abdominal Adipose Tissue and Thigh Muscle Volume and Hepatic Proton Density Fat Fraction: Repeatability and Accuracy of an MR Imaging-based, Semiautomated Analysis Method. Radiology 2017;283(2):438–449.

33. West J, Dahlqvist Leinhard O, Romu T, et al. Feasibility of MR-based body composition analysis in large scale population studies. PLoS One 2016;11(9):e0163332.

34. Nogami A, Yoneda M, Iwaki M, et al. Diagnostic comparison of vibration-controlled transient elastography and MRI techniques in overweight and obese patients with NAFLD. Sci Rep 2022;12(1):21925.

35. Caussy C, Chen J, Alquiraish MH, et al. Association Between Obesity and Discordance in Fibrosis Stage Determination by Magnetic Resonance vs Transient Elastography in Patients With Nonalcoholic Liver Disease. Clin Gastroenterol Hepatol 2018;16(12):1974–1982.e7.