

ORIGINAL ARTICLE

Influence of liver stiffness heterogeneity on staging fibrosis in patients with nonalcoholic fatty liver disease

Nobuyoshi Kawamura^{1,2}  | Kento Imajo^{1,2} | Kyle J. Kalutkiewicz³ | Koki Nagai^{1,2} | Michihiro Iwaki¹ | Takashi Kobayashi¹ | Asako Nogami¹  | Yasushi Honda¹ | Takaomi Kessoku¹ | Yuji Ogawa¹ | Takuma Higurashi¹ | Kunihiro Hosono¹ | Hirokazu Takahashi⁴  | Masato Yoneda¹ | Satoru Saito¹ | Shinichi Aishima⁵ | Hidenori Toyoda⁶ | Hideki Hayashi⁷ | Yoshio Sumida⁸ | Richard L. Ehman³ | Atsushi Nakajima¹

¹Department of Gastroenterology and Hepatology, Yokohama City University Graduate School of Medicine, Yokohama, Japan

²Department of Gastroenterology, Shin-yurigaoka General Hospital, Kawasaki City, Japan

³Department of Radiology, Mayo Clinic, Rochester, Minnesota, USA

⁴Division of Metabolism and Endocrinology, Faculty of Medicine, Saga University, Saga, Japan

⁵Department of Pathology and Microbiology, Faculty of Medicine, Saga University, Saga, Japan

⁶Department of Gastroenterology and Hepatology, Ogaki Municipal Hospital, Ogaki City, Japan

⁷Department of Gastroenterology and Hepatology, Gifu Municipal Hospital, Gifu City, Japan

⁸Department of Hepatology and Pancreatology, Aichi Medical University of Medicine, Nagakute, Japan

Correspondence

Atsushi Nakajima, Department of Gastroenterology and Hepatology, Yokohama City University Graduate School of Medicine, 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004, Japan.
Email: nakajima-ky@umin.ac.jp

Funding information

Supported by the "Step A" program of the Japan Science and Technology Agency (to J.S.T.) and Kiban-B, Shingakujuturyouiki, and, in part, by Grants-in-Aid from the Japanese Ministry of Health, Labour and Welfare (18K07637)

Abstract

Background and Aims: Despite that hepatic fibrosis often affects the liver globally, spatial distribution can be heterogeneous. This study aimed to investigate the effect of liver stiffness (LS) heterogeneity on concordance between MR elastography (MRE)-based fibrosis staging and biopsy staging in patients with NAFLD.

Approach and Results: We retrospectively evaluated data from 155 NAFLD patients who underwent liver biopsy and 3 Tesla MRE and undertook a retrospective validation study of 169 NAFLD patients at three hepatology centers. Heterogeneity of stiffness was assessed by measuring the range between minimum and maximum MRE-based LS measurement (LSM). Variability of LSM was defined as the stiffness range divided by the maximum stiffness value. The cohort was divided into two groups (homogenous or heterogeneous), according to whether variability was below or above the average for the training cohort. Based on histopathology and receiver operating characteristic

Abbreviations: 3T, 3 Tesla; ALT, alanine aminotransferase; AST, aspartate aminotransferase; AUROC, area under the receiver operating characteristic; BMI, body mass index; CLDs, chronic liver diseases; ELF, Enhanced Liver Fibrosis; FIB-4, Fibrosis-4; HA, hyaluronic acid; ICC, intraclass correlation coefficient; LB, liver biopsy; LF, liver fibrosis; LS, liver stiffness; LSM, liver stiffness measurement; MRE, MR elastography; NFS, NAFLD Fibrosis Score; ROC, receiver operating characteristic; ROI, region of interest.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Hepatology* published by Wiley Periodicals LLC on behalf of American Association for the Study of Liver Diseases

(ROC) analysis, optimum LSM thresholds were determined for MRE-based fibrosis staging of stage 4 (4.43, kPa; AUROC, 0.89) and stage ≥ 3 (3.93, kPa; AUROC, 0.89). In total, 53 had LSM above the threshold for stage 4. Within this group, 30 had a biopsy stage of <4 . In 86.7% of these discordant cases, variability of LSM was classified as heterogeneous. In MRE-based LSM stage ≥ 3 , 88.9% of discordant cases were classified as heterogeneous. Results of the validation cohort were similar to those of the training cohort.

Conclusions: Discordance between biopsy- and MRE-based fibrosis staging is associated with heterogeneity in LSM, as depicted with MRE.

INTRODUCTION

NAFLD is a leading cause of chronic liver disease (CLD), with a global increase in prevalence.^[1] Currently, the prevalence of NAFLD is estimated to be 25% in the general population, with 90% in obese persons and 60% in persons with type 2 diabetic mellitus having NAFLD. NAFLD ranges from benign nonalcoholic fatty liver with nonspecific inflammation to NASH with zone 3 hepatocellular ballooning (HB) or fibrosis, which is progressive and can lead to cirrhosis and HCC.^[2–5]

Therefore, early diagnosis of liver fibrosis (LF) and early interventions for NAFLD are important to reduce the risk of delayed complications.

Although liver biopsy (LB) remains the gold standard for the diagnosis of LF in CLDs,^[6] it has disadvantages, such as complications, high costs,^[7] and diagnostic variation, among observers.^[8–11] Sampling errors have also been identified as a weakness of this approach.^[12] In recent years, imaging-based methods such as MR elastography (MRE) have been developed as noninvasive methods for the evaluation of LF. The advantages of MRE are high diagnostic performance and the ability to assess the stiffness over a large volume of the liver.^[13–17] It has been suggested that spatial heterogeneity in the severity of LF may be an important factor when there is a discrepancy between biopsy- and MRE-based fibrosis staging.^[12] The goal of the present study was to investigate the potential relationship between MRE-based evidence of spatial heterogeneity in LF and discrepancies in biopsy staging in patients with NAFLD.

PATIENTS AND METHODS

Patients

We retrospectively evaluated the data of 155 NAFLD patients who underwent LB and 3 Tesla (3T) MRE between March 2014 and June 2021 at Yokohama City University Hospital. Furthermore, a total of 169 consecutive NAFLD patients who underwent LB and 3T MRE at Gifu Municipal Hospital, Ogaki Municipal Hospital,

and Shin-yurigaoka General Hospital from March 2018 through April 2021 were also enrolled in the study as a validation cohort. The study was conducted in accordance with the ethical principles of the Declaration of Helsinki (2013) and was approved by the Ethics Committee of Yokohama City University Hospital, and all patients provided written informed consent.

LB was performed within 6 months after MRE. Patients with a history of significant alcohol intake, chronic hepatitis, including viral and autoimmune hepatitis, and use of medications associated with fatty liver were excluded. Additionally, to avoid an incorrect assessment of liver stiffness measurement (LSM), patients were examined after fasting, and patients with known portal hypertension or passive hepatic congestion attributable to cardiac failure were also excluded (Figure S1).

Basic demographic data, including the age and sex of study participants, and relevant medical history, including diabetes mellitus, hypertension, or dyslipidemia, were acquired from medical records. Body mass index (BMI) was calculated as body mass in kilograms divided by the square of height in meters (kg/m^2). Aspartate aminotransferase (AST), alanine aminotransferase (ALT), γ -glutamyl transpeptidase (GGT), C-reactive protein, creatinine, fasting blood plasma glucose, fasting insulin, and hemoglobin A1c (HbA1c) were measured.

Histopathological evaluations

LB specimens were obtained from all patients using a 16-gauge needle biopsy kit (Bard MONOPTY; C.R. Bard, Inc., Murray Hill, NJ). For each patient, care was taken to collect a specimen sufficiently large for analysis. Adequate LB specimens were defined as those that were at least 15 mm in length and/or included at least six portal tracts. Central pathology determination was used for histological evaluation. LB slides, stained with hematoxylin and eosin as well as Masson stain, were evaluated by a single experienced central pathologist (S.A.) who specializes in liver histopathology. In addition, the intra- and interobserver variations for pathological LF

stages were analyzed by rereading by a central pathologist (S.A., reader 1) and hepatologist (reader 2).

Steatosis, lobular inflammation, and ballooned hepatocytes were classified as follows. Steatosis affecting <5%, 5%–33%, >33%–66%, and >66% of hepatocytes was classified as grades 0, 1, 2, and 3, respectively. Lobular inflammation was graded according to the number of inflammatory foci per field of view (FOV) at a magnification of 200×, with 0, <2, 2–4, and >4 foci per field classified as grades 0, 1, 2, and 3, respectively. HB involving no, few, and many cells was classified as grades 0, 1, and 2, respectively. Fibrosis severity was scored as described.^[18]

Serum biomarker

Serum hyaluronic acid (HA) was measured as a biomarker for LF, using the latex agglutination immunoassay method (Mitsubishi Chemical, Tokyo, Japan).^[19] The Enhanced Liver Fibrosis (ELF) score, considering tissue inhibitor of metalloproteinases 1, amino-terminal propeptide of type III procollagen, and HA, was also measured as a biomarker for LF.^[20]

Scoring systems

Based on a review of the literature, the following scores were calculated for each patient to evaluate LF: Fibrosis-4 (FIB-4) index $[\text{age (year)} \times \text{AST (IU/L)} / (\text{platelet count} (\times 10^9/\text{L}) \times \sqrt{\text{ALT (IU/L)}})]$ ^[21]; NFS (NAFLD Fibrosis Score $[-1.675 + 0.037 \times \text{age (year)} + 0.094 \times \text{BMI (kg/m}^2) + 1.13 \times \text{impaired fasting glycemia/diabetes (present = 1, absent = 0)} + 0.99 \times \text{AST (IU/L)/ALT (IU/L)} - 0.013 \times \text{platelet count} (\times 10^9/\text{L}) - 0.66 \times \text{albumin (g/dL)}]$.^[22]

MRE

All eligible patients underwent hepatic MRE using 3T imagers (GE Healthcare, Milwaukee, WI) installed in our hospital. MRE was performed after fasting for 12 h. Continuous longitudinal mechanical waves (60 Hz) were generated using a passive acoustic driver placed against the anterior chest wall. According to previous methods,^[13] a two-dimensional spin-echo planar MRE sequence was used to acquire wave images in transverse sections with the following parameters: repetition time ms/echo time ms, 50/23; continuous sinusoidal vibration, 60 Hz; FOV, 32–42 cm; matrix size, 256 × 64; flip angle, 30 degrees; section thickness, 10 mm; four evenly spaced phase offsets; and four pairs of 60-Hz trapezoidal motion-encoding gradients with zero- and first-moment nulling along the through-plane direction. All processing steps were applied automatically, without manual intervention, to yield quantitative images of

tissue shear stiffness in kilopascals. On each section of the MR magnitude image from the MRE acquisition, regions of interest (ROIs) were drawn to include only the parenchyma of the right lobe, avoiding the edges of the liver and large blood vessels. The mean of measurements in four slices was used for analysis.^[15] If there were no liver parenchyma that could be measured by elastograms using reliability maps, the study was considered invalid. ROIs also excluded regions where the phase signal-to-noise ratio (the ratio of the wave amplitude to the noise in the wave images) was <5. The LSM obtained at the time of examination was entered into the database and extracted for this study.

Analysis of liver stiffness heterogeneity

Variability in liver stiffness (LS) was used as an indicator of LS heterogeneity in MRE. The lowest and highest sites of LS were measured using 1-cm² ROIs. Variability was defined as the difference between the maximum and minimum LS values, divided by the maximum value, and is expressed in percentage. In addition, the amount of overlap of the stiffness ranges below cut-off values for each fibrosis stage, such as stage 4, fibrosis stage ≥3, fibrosis stage ≥2, and fibrosis stage ≥1, was calculated for each case. Overlap was defined as (cut-off value – the minimum ROI stiffness value) / (maximum – minimum ROI stiffness value) and is expressed in percentage.

Statistical analysis

Continuous variables were summarized as means and SDs. Because many variables were not normally distributed, the Kruskal-Wallis test was used for comparisons of more than two independent groups, and *p* values <0.05 were considered statistically significant. Cut-off values were calculated by using the Youden index. All statistical analyses were performed using JMP software (Pro 15; SAS Institute Inc., Cary, NC). We used the intraclass correlation coefficient (ICC) to assess the intra- and interobserver variation for pathological LF stages in the training cohort. ICCs were obtained using a two-way random-effects model for absolute agreement and a single rater.^[23]

RESULTS

Patient characteristics

In this retrospective study, 155 patients with LB-proven NAFLD were enrolled in the training cohort and 169 patients with LB-proven NAFLD were enrolled in the validation cohort. Principal features, laboratory characteristics, pathological LF stages, and other pathological findings

TABLE 1 Clinical, serological, and histological characteristics of patients

Characteristic	Training Cohort	Validation Cohort
<i>n</i>	155	169
Age, years, mean ± SD	59.3 ± 13.6	61.8 ± 12.5
Sex, male/female	86/69	76/93
BMI, kg/m ² , mean ± SD	29.91 ± 5.73	27.67 ± 4.85
Platelets, /10 ⁴ μl, mean ± SD	19.60 ± 6.88	20.82 ± 6.35
AST, IU/L, mean ± SD	51.3 ± 27.3	51.0 ± 28.1
ALT, IU/L, mean ± SD	65.2 ± 42.0	64.3 ± 43.4
GGT, mean ± SD	86.2 ± 93.7	75.1 ± 70.0
Fasting blood glucose, mg/dL, mean ± SD	116.8 ± 29.2	113.5 ± 27.4
HbA1c, mean ± SD	6.47 ± 1.09	6.21 ± 0.87
Diabetes mellitus, %	87 (56.1)	61 (36.1)
Hypertension, %	76 (49.0)	92 (54.4)
Dyslipidemia, %	102 (65.8)	85 (50.3)
Steatosis grade, <i>n</i>		
0	9	14
1	72	118
2	50	34
3	24	3
Lobular inflammation, <i>n</i>		
0	7	4
1	115	88
2	32	76
3	1	1
Liver cell ballooning, <i>n</i>		
0	65	30
1	62	72
2	28	65
NAS, <i>n</i>		
0/1/2/3/4/5/6/7/8	3/6/32/37/41/24/10/1/1	1/7/16/42/54/34/14/1
Fibrosis, <i>n</i>		
0	12	9
1	35	55
2	42	45
3	43	43
4	23	17

Abbreviation: NAS, NAFLD Activity Score.

are summarized in Table 1. Intra- and interobserver variations in pathological fibrosis stages are demonstrated in Table S1. ICCs for both were good or excellent.

Diagnostic accuracy of serum biomarkers and scores for LF detection in patients with NAFLD

HA and ELF were measured as serum biomarkers of fibrosis. Additionally, the FIB-4 index and NFS were used for scoring fibrosis. The diagnostic performance of these four parameters as well as their areas under the receiver operating characteristic curves (AUROCs) and their optimal cut-off levels for diagnosing LF stages ≥2, ≥3, and 4 are shown in Table S2.

Diagnostic accuracy of MRE for LF detection in patients with NAFLD

LSM was determined using MRE in patients with NAFLD, to assess the stage of LF. Mean LSM values for MRE were 2.40, 2.78, 3.58, 4.83, and 5.67 kPa for patients with biopsy-determined stages 0, 1, 2, 3, and 4, respectively (Figure S2). Results of these analyses revealed step-wise increases in the LSM obtained using MRE with increasing histological severity of hepatic fibrosis ($p < 0.0001$, Kruskal-Wallis test). To investigate the diagnostic accuracy of the LSM obtained using MRE for LF in this cohort, we performed receiver operating characteristic (ROC) analysis, with AUROC and cut-off values shown in Table S3 and ROC curves in Figure S3.

Heterogeneity of LS

Distribution of stiffness variability is shown in Figure 1. Variability according to the pathological fibrosis stage is shown in Figure S4. The mean variability value was 30.59%. A high rate of variability corresponds to a more heterogeneous distribution of LSM, whereas a low rate of variability reflects a more homogeneous distribution of LSM. The mean value of variability was used to divide the MRE exams into two groups: relatively heterogeneous and relatively homogeneous. Figure 2 shows representative MRE images of higher variability and lower variability in LS. Additionally, patient characteristics of heterogeneous and homogeneous types were compared (Table S4).

Discordance between MRE-based and pathological fibrosis staging with pathological underestimation and relationship to LS heterogeneity

Based on the ROC-defined optimum cut-off value of 4.43 kPa, 53 patients were classified as having stage 4 fibrosis by MRE in the training cohort (Table 2). Histopathological staging was discordant (showing less than stage 4 fibrosis) in 30 patients in this group. Based on the cut-off value of 4.43 kPa, 32 patients were

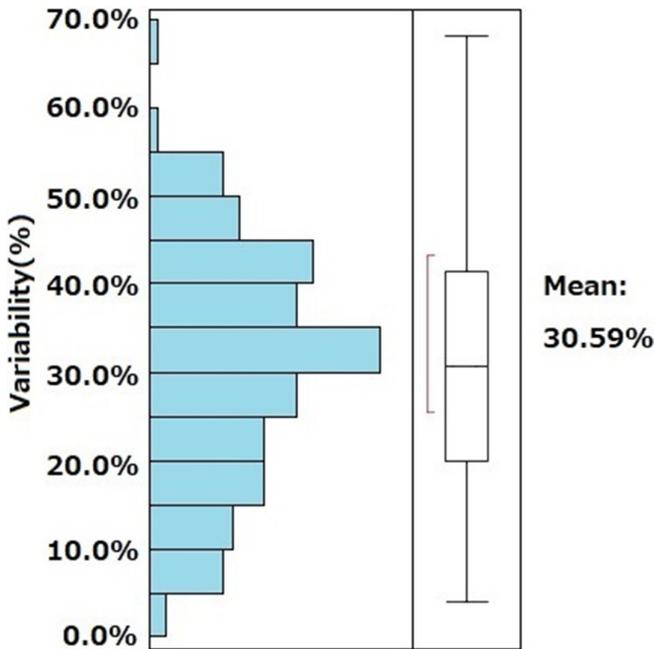
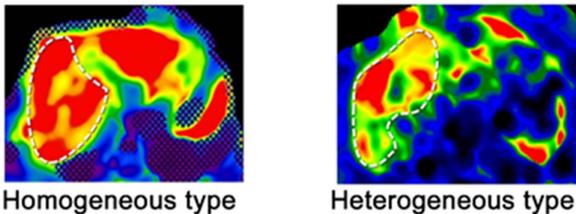


FIGURE 1 Histogram illustrating the distribution of stiffness variability in the training study cohort, where variability was defined as the difference between maximum and minimum LS values (measured in 1-cm² ROIs), divided by the maximum value and expressed in percentage

(A) Conventional MRE elastogram



(B) 3D images of MRE elastogram

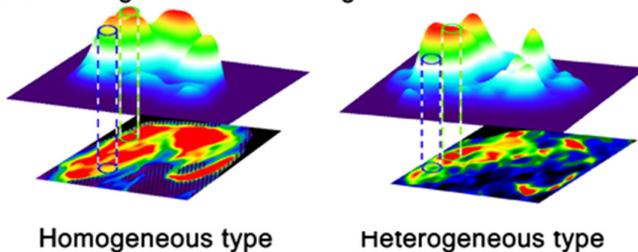


FIGURE 2 (A) MRE images demonstrating more heterogeneous (right) and less heterogeneous (left) LS patterns. The homogeneous type with advanced LS appears as a uniformly red signal on conventional MRE images. However, in the heterogeneous type, only partial areas show the red signal, indicating advanced LS. (B) Conventional MRE images were converted into three-dimensional (3D) images by representing the high and low LS regions as contour lines, making it easier to assess the areas with highest and lowest LS visually

classified as having stage 4 fibrosis by MRE in the validation cohort (Table 2). Histopathological staging was discordant (showing less than stage 4 fibrosis) in 15

patients in this cohort. Among the discordant cases, 86.7% or 80.0% had LS patterns classified as heterogeneous, whereas in the concordant cases, only 43.5% or 12.5% had LS patterns classified as heterogeneous in the training cohort or validation cohort (Table 3). Nevertheless, the mean values of laboratory biomarkers HA, ELF, FIB-4 index, and NFS in the group of 30 discordant cases exceeded the cut-off values for stage 4 fibrosis (Table 3).

A total of 67 patients were classified as having stage ≥ 3 fibrosis by MRE using the ROC-defined optimum cut-off value of 3.93 kPa in the training cohort (Table 2). Histopathological staging was discordant in only 9 patients. Based on the cut-off value of 3.93 kPa, a total of 42 patients were classified as having stage ≥ 3 fibrosis by MRE in the validation cohort (Table 2). Histopathological staging was discordant in only 10 patients. Among the discordant cases, 88.9% or 80.0% had LS patterns classified as heterogeneous, whereas in the concordant cases only 62.1% or 34.4% had patterns classified as heterogeneous in the training cohort or validation cohort (Table 4). Mean values of HA, FIB-4 index, and NFS in the group of 9 discordant cases exceeded the cut-off values for stage 3 fibrosis. Mean values of ELF were slightly below the cut-off values, but almost the same for stage 3 fibrosis.

To further evaluate the relationship between concordance and stiffness heterogeneity, a two-sample (variability and range overlap) Hotelling T-squared test statistic was computed along with its associated p value. Distribution of concordant and discordant cases based on two metrics (variability and overlap) and its associated p value by the two-sample Hotelling T-squared test analysis for the training and validation cohort are shown in Figure 3A,B (left column) and Table 5. Distribution of concordant and discordant cases was clearly separated for MRE-based LSM stage 4 and ≥ 3 (Figure 3A,B, left column). The p value showed a strong statistically significant difference between the discordant group and concordant group in MRE-based LSM stages 4 and ≥ 3 (Table 5). In the combined cohort, the significance of the difference was even greater. Distribution of concordant and discordant cases by the two-sample Hotelling T-squared test analysis for combined cohort is shown in Figure S5 (left column).

Similar results were obtained in lower fibrosis stages, such as MRE-based LSM stage 2, compared with higher fibrosis stages. All of the discordant cases were of the heterogeneous type in the training cohort (Table S5). In the validation cohort, 87.5% of discordant cases were classified as the heterogeneous type. However, based on the p value by the two-sample Hotelling T-squared test analysis, there was a trend, but it did not reach significance between the discordant and concordant groups even in the combined cohort (Figure 3A,B and Figure S5; Table S5). *Italic indicates significant P values <0.05.*

TABLE 2 Discordance between MRE-based LSM stage and pathologically advanced fibrosis stage

	Training Cohort			Validation Cohort		
	MRE-Based LSM			MRE-Based LSM		
	Stage 4	Stage ≥ 3	Stage ≥ 2	Stage 4	Stage ≥ 3	Stage ≥ 2
	(Cutoff, 4.43 kPa)	(Cutoff, 3.93 kPa)	(Cutoff, 3.00 kPa)	(Cutoff, 4.43 kPa)	(Cutoff, 3.93 kPa)	(Cutoff, 3.00 kPa)
<i>n</i>	53	67	106	32	42	85
Pathological fibrosis stage						
0	0	0	0	0	0	0
1	0	1	6	0	1	8
2	5	8	34	5	9	24
3	25	35	43	10	15	36
4	23	23	23	17	17	17

TABLE 3 Association between MRE-based high LSM (stage 4) and pathological fibrosis stage in terms of variability and fibrosis markers in the training and validation cohorts

	MRE-Based LSM Stage 4			
	Training Cohort		Validation Cohort	
	Discordant	Concordant	Discordant	Concordant
<i>n</i>	30	23	15	17
Heterogeneous, %	86.7 (<i>n</i> = 26)	43.5 (<i>n</i> = 10)	80.0 (<i>n</i> = 12)	12.5 (<i>n</i> = 2)
HA, mean \pm SD	200.0 \pm 210.4	214.1 \pm 247.4	N/A	N/A
ELF, mean \pm SD	11.18 \pm 0.96	11.28 \pm 0.94	N/A	N/A
FIB-4 index, mean \pm SD	3.81 \pm 1.92	3.88 \pm 1.89	3.60 \pm 1.32	3.46 \pm 1.97
NFS, mean \pm SD	1.16 \pm 1.29	1.04 \pm 1.16	-0.11 \pm 1.72	0.22 \pm 1.22

Note: Cut-off values of HA, ELF, FIB-4 index, and NFS for fibrosis stage 4 were 83, 10.77, 2.13, and -0.59, respectively.

Abbreviation: N/A, not applicable.

TABLE 4 Association between MRE-based high LSM (stage ≥ 3) and pathological stage in terms of variability and fibrosis markers

	MRE-Based LSM Stage ≥ 3			
	Training Cohort		Validation Cohort	
	Discordant	Concordant	Discordant	Concordant
<i>n</i>	9	58	10	32
Heterogeneous, %	88.9 (<i>n</i> = 8)	62.1 (<i>n</i> = 36)	80.0 (<i>n</i> = 8)	34.4 (<i>n</i> = 11)
HA, mean \pm SD	117.8 \pm 86.7	195.7 \pm 217.2	N/A	N/A
ELF, mean \pm SD	10.70 \pm 0.81	11.19 \pm 0.91	N/A	N/A
FIB-4 index, mean \pm SD	3.32 \pm 1.33	3.61 \pm 1.98	2.55 \pm 0.68	3.58 \pm 1.66
NFS, mean \pm SD	0.76 \pm 1.73	0.75 \pm 1.46	-0.88 \pm 1.20	0.26 \pm 1.27

Note: Cut-off values of HA, ELF, FIB-4 index, and NFS for fibrosis stage ≥ 3 were 69, 10.73, 2.09, and -1.06, respectively.

Abbreviation: N/A, not applicable.

Discordance between MRE-based and pathological fibrosis staging with MRE downstimation and relationship to LS heterogeneity

There were no cases with MRE-based LSM stage < 3 and histopathological fibrosis stage 4.

A total of 66 patients had histopathological fibrosis stage ≥ 3 in the training cohort. Among this group, 8 patients were classified as having stage ≤ 2 by MRE using a cut-off value of 3.93 kPa (Table S6). In the validation cohort, a total of 60 patients had histopathological fibrosis stage ≥ 3 . Based on the ROC-defined optimum cut-off value of 3.93 kPa, 28 patients were classified as

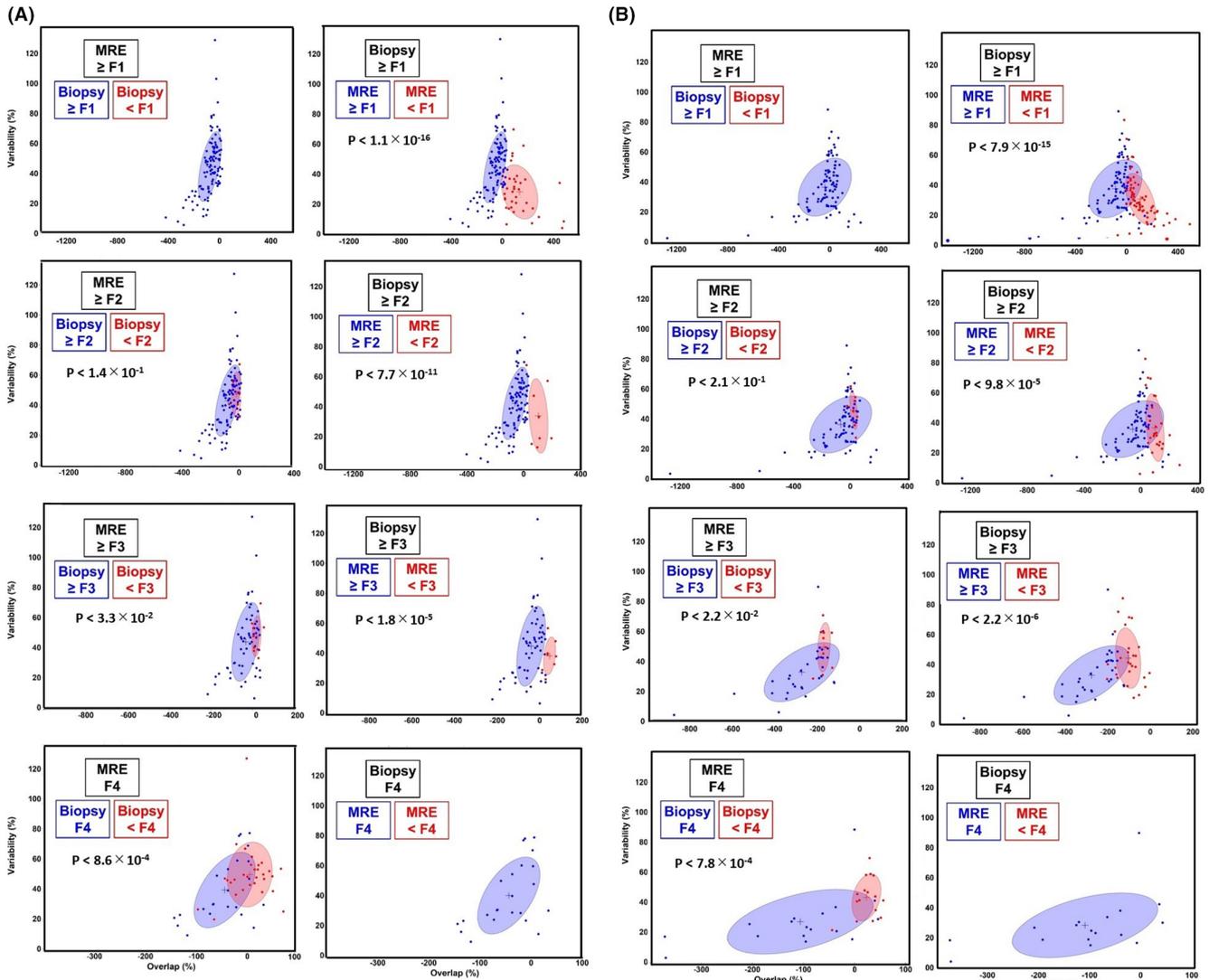


FIGURE 3 (A) Left column: graph showing two metrics of LS heterogeneity (variability and overlap, as defined in the text) for each MRE-based LSM stage in the training cohort. Cases in which histopathology staging was lower than MRE-based LSM staging are shown in red. Probability distributions of concordant and discordant cases were significantly different, except MRE-based LSM stage 2. Discordant biopsy results were associated with higher metrics of LS heterogeneity. Right column: similar graph for each histopathological fibrosis stage in the training cohort. Discordant cases, in which MRE-based LSM stage was lower than pathological fibrosis stage, are shown in red. Probability distributions of concordant and discordant cases were significantly different. (B) Similar graph to (A) in the validation cohort

having stage ≤ 2 . In the training cohort, mean values of fibrosis markers and scores in discordant cases supported fibrosis stage ≤ 2 , except for the NFS (Table S6).

Similar results were obtained for patients with lower histopathological fibrosis, such as stage ≥ 2 and stage ≥ 1 , in the training and validation cohorts. There were 8 or 28 discordant cases with MRE-based LSM stage ≤ 1 in the training or validation cohort, respectively. There were 35 or 67 discordant cases with MRE-based LSM stage 0 in the training or validation cohort, respectively (Table S6).

To evaluate the relationship between concordance and stiffness heterogeneity, a two-sample (variability and range overlap) Hotelling T-squared test statistic was computed along with its associated p value.

Distribution of concordant and discordant cases based on two metrics (variability and overlap) and its associated p value by the two-sample Hotelling T-squared test analysis for the training and validation cohort are shown in Figure 3A,B (right column) and Table 5. Distribution of concordant and discordant cases was clearly separated for each pathological fibrosis stage (Figure 3A,B, right column). The p value showed a strong statistically significant difference between the discordant group and concordant group in all fibrosis stages (Table 5). In the combined cohort, the significance of the difference was even greater. Distribution of concordant and discordant cases by the two-sample Hotelling T-squared test analysis for the combined cohort is shown in Figure S5 (right column).

TABLE 5 A two-sample (variability and overlap) Hotelling T-squared test *p* values for each fibrosis stage based on MRE or biopsy in the training, validation, and combined cohorts

	Training Cohort	Validation Cohort	Combined Cohort
MRE, stage			
≥2	0.137	0.205	0.12
≥3	<i>0.033</i>	<i>0.022</i>	<i>0.0014</i>
≥4	<i>8.60 × 10⁻⁴</i>	<i>7.76 × 10⁻⁴</i>	<i>1.7 × 10⁻⁶</i>
Biopsy, stage			
≥1	<i>1.11 × 10⁻¹⁶</i>	<i>7.88 × 10⁻¹⁵</i>	<i>1.1 × 10⁻¹⁶</i>
≥2	<i>7.66 × 10⁻¹¹</i>	<i>9.84 × 10⁻⁵</i>	<i>2.0 × 10⁻¹²</i>
≥3	<i>1.84 × 10⁻⁵</i>	<i>2.15 × 10⁻⁶</i>	<i>1.1 × 10⁻⁶</i>

Italic indicates significant *p* values < 0.05.

DISCUSSION

Multiple published studies have provided evidence that MRE has high diagnostic performance for detecting and staging LF in NAFLD patients.^[13–17] Similarly, MRE showed high diagnostic accuracy for staging LF in this study as reported.^[24,25] The diagnostic performance demonstrated in many studies approaches the limit that can be demonstrated, given the known limitations of reliability of biopsy-based staging attributable to subjective interpretation and sampling error. However, one report suggested that MRI-based proton density fat fraction is superior to LB for the assessment of hepatic steatosis rather than fibrosis, because of sampling error.^[26]

Although sampling error is often acknowledged as a limitation of biopsy, the evaluation of whether discordance with MRE-based fibrosis staging is associated with spatial heterogeneity of stiffness as depicted by MRE is a strength of this study.

In an LB, only 1/50,000th of the whole liver tissue is evaluated, raising a substantial possibility of sampling error. In a previous study, Ratziu et al.^[12] performed two percutaneous LBs in each of 51 patients with NAFLD and reported on sampling variability. They showed that in 41% of patients, the two specimens demonstrated different stages of fibrosis. They speculated that the discordance between specimens is likely attributable to heterogeneous distribution of pathological features throughout the liver. In another study with a different etiology of LF, Regev et al.^[27] reported that biopsies from the right and left lobes of the liver differed by more than one pathological stage in 30% of patients in HCV patients.

Despite not discussing the discordance between pathological fibrosis stage and MRE-based LSM, Caussy et al.^[28] reported discordance between fibrosis stages obtained by MRE and vibration-controlled transient elastography (VCTE) in ~45% of patients with NAFLD. They showed that the higher the BMI of the patient, the greater the discrepancy. They speculated

that the higher BMI resulted in a higher skin-to-capsule distance. Additionally, they also proposed that VCTE assesses LS in distinct locations, whereas MRE may reduce sampling error given that it assesses the whole liver. Bedossa et al.^[29] argued that sampling variability in HCV patients is affected by length of the biopsy specimen.

Based on these and other publications, there is a reasonable expectation that heterogeneity in the severity of LF in CLDs, such as NAFLD and HCV, may contribute to discordance between biopsy-based fibrosis staging and noninvasive staging methods that assess a larger volume of the liver—such as MRE. Presence of marked heterogeneity in LS is often observed with MRE, as illustrated in Figure 2. The results of this study provide support for the hypothesis that discordant results between MRE- and biopsy-based fibrosis staging can be the result of sampling error. In patients with global LSMs consistent with either stage 4 or stage ≥3 fibrosis, discrepant biopsy results were more common in the presence of marked LS heterogeneity. In patients with LSM consistent with stage ≥2 fibrosis, there was a tendency for LS classified into the heterogeneous pattern in patients with discrepant biopsy results, although the analysis using a two-sample (variability and overlap) Hotelling T-squared test did not reveal significant differences. This result was also observed in the validation cohort and was postulated to be influenced by the small sample size; thus, it was also examined in the combined cohort. The results did indeed show a smaller *p* value, despite still not being statistically significant. Thus, it is possible that a statistically significant difference could be obtained by increasing the sample size. However, in cases where the MRE results are lower than the pathological fibrosis stage, the presence of marked LS heterogeneity was more common. Similar results were obtained in the validation cohort. In other words, the involvement of heterogeneity in LS was suspected even in discordant cases in which biopsy was not only underestimated, but also overestimated (MRE downestimation). These results support the hypothesis that LS heterogeneity is responsible for the discordance with biopsy results. In this study, the validity of MRE-based fibrosis staging was supported by concordance of other fibrosis markers, such as HA, ELF, FIB-4 index, and the NFS.

Clinically, it may be important to recognize that an LB is more likely to result in understaging because of sampling error in patients with high spatial heterogeneity in LS. Hence, biopsy staging may need to be interpreted with more caution when high heterogeneity is present in clinical trials or to determine treatment efficacy.

The limitations of this study were that it was a retrospective study with the number of cases limited to 324 (155 patients in the training cohort and 169 in the validation cohort). Another limitation was that we were unable to obtain supportive data, such as fibrosis

markers, for the validation cohort because of the lack of serum samples. Moreover, the method of obtaining the ROI was subjective, and the exact location of the LB was not shown in the MRE.

In this study, we provide evidence that heterogeneity in the spatial severity of LF, as reflected in LS heterogeneity, may be instrumental when there is discordance between MRE-based staging and pathological staging in patients with NAFLD. In particular, patients with spatially heterogeneous LS may be more likely to have erroneously over-/downestimated pathological fibrosis staging because of sampling error if an LB is used as the gold standard. Further research is needed to explore the implications of these findings in patients with NAFLD.

CONFLICT OF INTEREST

Nothing to report.

AUTHOR CONTRIBUTIONS

Conception and design of the study: Nobuyoshi Kawamura, Kento Imajo, Hirokazu Takahashi, Masato Yoneda, Richard L. Ehman, and Atsushi Nakajima. *Data collection:* Koki Nagai, Michihiro Iwaki, Takashi Kobayashi, Asako Nogami, Yasushi Honda, Takaomi Kessoku, Yuji Ogawa, Hidenori Toyoda, Hideki Hayashi, Yoshio Sumida, and Satoru Saito. *Operators of elastographies:* Kento Imajo (an elastography expert). *Data analyses:* Nobuyoshi Kawamura, Kento Imajo, Atsushi Nakajima, Kyle J. Kalutkiewicz, and Richard L. Ehman. *Contribution of reagents/materials/analytical tools:* Nobuyoshi Kawamura, Kento Imajo, Masato Yoneda, and Atsushi Nakajima. *Pathological findings:* Shinichi Aishima. *Manuscript preparation:* Nobuyoshi Kawamura, Kento Imajo, Takuma Higurashi, Kunihiro Hosono, Masato Yoneda, Atsushi Nakajima, and Richard L. Ehman.

ORCID

Nobuyoshi Kawamura  <https://orcid.org/0000-0002-4201-3534>

Asako Nogami  <https://orcid.org/0000-0002-6923-365X>

Hirokazu Takahashi  <https://orcid.org/0000-0002-2374-2889>

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;64:73–84.
2. Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Non-alcoholic fatty liver disease: a spectrum of clinical and pathologic severity. *Gastroenterology*. 1999;116:1413–9.
3. Day CP, Saksena S. Non-alcoholic steatohepatitis: definitions and pathogenesis. *J Gastroenterol Hepatol*. 2002;17(Suppl. 3):S377–84.
4. Harrison SA, Torgerson S, Hayashi PH. The natural history of nonalcoholic fatty liver disease: a clinical histopathological study. *Am J Gastroenterol*. 2003;98:2042–7.
5. Singh S, Allen AM, Wang Z, Prokop LJ, Murad MH, Loomba R. Fibrosis progression in nonalcoholic fatty liver vs non-alcoholic steatohepatitis: a systematic review and meta-analysis of paired-biopsy studies. *Clin Gastroenterol Hepatol*. 2015;13:643–54.e1.
6. Angulo P. Nonalcoholic fatty liver disease. *N Engl J Med*. 2002;346:1221–31.
7. Cadranel JF. Good clinical practice guidelines for fine needle aspiration biopsy of the liver: past, present and future. *Gastroenterol Clin Biol*. 2002;26:823–4.
8. Janiec DJ, Jacobson ER, Freeth A, Spaulding L, Blaszyk H. Histologic variation of grade and stage of non-alcoholic fatty liver disease in liver biopsies. *Obes Surg*. 2005;15:497–501.
9. Younossi ZM, Gramlich T, Liu YC, Matteoni C, Petrelli M, Goldblum J, et al. Nonalcoholic fatty liver disease: assessment of variability in pathologic interpretations. *Mod Pathol*. 1998;11:560–5.
10. Gawrieh S, Knoedler DM, Saeian K, Wallace JR, Komorowski RA. Effects of interventions on intra- and interobserver agreement on interpretation of nonalcoholic fatty liver disease histology. *Ann Diagn Pathol*. 2011;15:19–24.
11. Kuwashiro T, Takahashi H, Hyogo H, Ogawa Y, Imajo K, Yoneda M, et al. Discordant pathological diagnosis of non-alcoholic fatty liver disease: a prospective multicenter study. *JGH Open*. 2020;4:497–502.
12. Ratziu V, Charlotte F, Heurtier A, Gombert S, Giral P, Bruckert E, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology*. 2005;128:1898–906.
13. Imajo K, Kessoku T, Honda Y, Tomeno W, Ogawa Y, Mawatari H, 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.e7.
14. Huwart L, Sempoux C, Vicaute E, Salameh N, Annet L, Danse E, et al. Magnetic resonance elastography for the noninvasive staging of liver fibrosis. *Gastroenterology*. 2008;135:32–40.
15. Loomba R, Wolfson T, Ang B, Hooker J, Behling C, Peterson M, et al. Magnetic resonance elastography predicts advanced fibrosis in patients with nonalcoholic fatty liver disease: a prospective study. *Hepatology*. 2014;60:1920–8.
16. Chen J, Talwalkar JA, Yin M, Glaser KJ, Sanderson SO, Ehman RL. Early detection of nonalcoholic steatohepatitis in patients with nonalcoholic fatty liver disease by using MR elastography. *Radiology*. 2011;259:749–56.
17. Kim D, Kim WR, Talwalkar JA, Kim HJ, Ehman RL. Advanced fibrosis in nonalcoholic fatty liver disease: noninvasive assessment with MR elastography. *Radiology*. 2013;268:411–9.
18. Brunt EM. Nonalcoholic steatohepatitis: definition and pathology. *Semin Liver Dis*. 2001;21:3–16.
19. Suzuki A, Angulo P, Lymp J, Li D, Satomura S, Lindor K. Hyaluronic acid, an accurate serum marker for severe hepatic fibrosis in patients with non-alcoholic fatty liver disease. *Liver Int*. 2005;25:779–86.
20. Lichtigthagen R, Pietsch D, Bantel H, Manns MP, Brand K, Bahr MJ. The enhanced liver fibrosis (ELF) score: normal values, influence factors and proposed cut-off values. *J Hepatol*. 2013;59:236–42.
21. Sumida Y, Yoneda M, Hyogo H, Itoh Y, Ono M, Fujii H, et al. Validation of the FIB4 index in a Japanese nonalcoholic fatty liver disease population. *BMC Gastroenterol*. 2012;12:2.
22. Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, Farrell GC, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology*. 2007;45:846–54.
23. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1:30–46.
24. Hsu C, Caussy C, Imajo K, Chen J, Singh S, Kaulback K, et al. Magnetic resonance vs transient elastography analysis

- of patients with nonalcoholic fatty liver disease: a systematic review and pooled analysis of individual participants. *Clin Gastroenterol Hepatol*. 2019;17:630–7.e8.
25. Kennedy P, Wagner M, Castéra L, Hong CW, Johnson CL, Sirlin CB, et al. Quantitative elastography methods in liver disease: current evidence and future directions. *Radiology*. 2018;286:738–63.
 26. Nouredin M, Lam J, Peterson MR, Middleton M, Hamilton G, Le TA, et al. Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials. *Hepatology*. 2013;58:1930–40.
 27. Regev A, Berho M, Jeffers LJ, Milikowski C, Molina EG, Pyrsopoulos NT, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am J Gastroenterol*. 2002;97:2614–8.
 28. Caussy C, Chen J, Alqiraish MH, Cepin S, Nguyen P, Hernandez C, 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;16:1974–82.e7.
 29. Bedossa P, Dargère D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology*. 2003;38:1449–57.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Kawamura N, Imajo K, Kalutkiewicz KJ, Nagai K, Iwaki M, Kobayashi T, et al. Influence of liver stiffness heterogeneity on staging fibrosis in patients with nonalcoholic fatty liver disease. *Hepatology*. 2022;76:186–195.

<https://doi.org/10.1002/hep.32302>