



Prediction of nonalcoholic fatty liver disease (NAFLD) activity score (NAS) with multiparametric hepatic magnetic resonance imaging and elastography

Ziying Yin¹ · Matthew C. Murphy¹ · Jiahui Li¹ · Kevin J. Glaser¹ · Amy S. Mauer² · Taofic Mounajjed³ · Terry M. Therneau⁴ · Heshan Liu⁴ · Harmeet Malhi² · Armando Manduca^{1,5} · Richard L. Ehman¹ · Meng Yin¹

Received: 6 September 2018 / Revised: 17 January 2019 / Accepted: 6 February 2019 / Published online: 18 March 2019

© European Society of Radiology 2019

Abstract

Objectives To investigate the use of MR elastography (MRE)–derived mechanical properties (shear stiffness ($|G^*|$) and loss modulus (G'')) and MRI-derived fat fraction (FF) to predict the nonalcoholic fatty liver disease (NAFLD) activity score (NAS) in a NAFLD mouse model.

Methods Eighty-nine male mice were studied, including 64 training and 25 independent testing animals. An MRI/MRE exam and histologic evaluation were performed. Pairwise, nonparametric comparisons and multivariate analyses were used to evaluate the relationships between the three imaging parameters (FF, $|G^*|$, and G'') and histologic features. A virtual NAS score (vNAS) was generated by combining three imaging parameters with an ordinal logistic model (OLM) and a generalized linear model (GLM). The prediction accuracy was evaluated by ROC analyses.

Results The combination of FF, $|G^*|$, and G'' predicted $NAS > 1$ with excellent accuracy in both training and testing sets (AUROC > 0.84). OLM and GLM predictive models misclassified 3/54 and 6/54 mice in the training, and 1/25 and 1/25 in the testing cohort respectively, in distinguishing between “not-NASH” and “definite-NASH.” “Borderline-NASH” prediction was poorer in the training set, and no borderline-NASH mice were available in the testing set.

Conclusion This preliminary study shows that multiparametric MRI/MRE can be used to accurately predict the NAS score in a NAFLD animal model, representing a promising alternative to liver biopsy for assessing NASH severity and treatment response.

Key Points

- MRE-derived liver stiffness and loss modulus and MRI-assessed fat fraction can be used to predict NAFLD activity score (NAS) in our preclinical mouse model (AUROC > 0.84 for all NAS levels greater than 1).
- The overall agreement between the histological-determined NASH diagnosis and the imaging-predicted NASH diagnosis is 80–92%.
- The multiparametric hepatic MRI/MRE has great potential for noninvasively assessing liver disease severity and treatment efficacy.

Keywords Magnetic resonance imaging · Elastography · Nonalcoholic fatty liver disease · Nonalcoholic steatohepatitis

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00330-019-06076-0>) contains supplementary material, which is available to authorized users.

✉ Meng Yin
yin.meng@mayo.edu

¹ Department of Radiology, Mayo Clinic, 200 First St SW, Rochester, MN 55905, USA

² Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA

³ Division of Anatomic Pathology, Mayo Clinic, Rochester, MN, USA

⁴ Department of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA

⁵ Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, MN, USA

Abbreviations

AUROC	Area under receiver operating characteristic curve
FF	Fat fraction
FOV	Field of view
GLM	Generalized linear model
H&E	Hematoxylin-eosin
IQR	Interquartile range
MRE	Magnetic resonance elastography
NAFLD	Nonalcoholic fatty liver disease
NAS	NAFLD activity score
NASH	Nonalcoholic steatohepatitis
OLM	Ordinal logistic model
ROC	Receiver operating characteristic
ROI	Region of interest
TE	Echo time
TR	Repetition time
vNAS	Virtual NAS

Introduction

Nonalcoholic fatty liver disease (NAFLD) is significant public health problems worldwide [1–3], which ranges from simple nonalcoholic fatty liver to nonalcoholic steatohepatitis (NASH) with advanced stages of fibrosis. NASH patients may progress to cirrhosis and confer an increased risk of developing hepatocarcinoma [4–6]. Currently, the translation of basic research into improved therapeutics and management of NAFLD patients is still poor as the invasive biopsy has been the bedrock for screening, longitudinal monitoring, or evaluating treatment response. Therefore, development of noninvasive imaging biomarkers of liver injuries is critically important for clinical management and for a better understanding of the disease progression.

Many investigations have produced potential imaging biomarkers for noninvasively quantifying specific histologic features of NAFLD. For instances, MRI offers chemical-shift imaging and spectroscopic methods for quantifying steatosis (fat fraction, or FF) with relatively high accuracy without any invasive procedures or radiation exposure, compared with invasive liver biopsy, qualitative ultrasound, and semi-quantitative CT [7–10]. MRI-measured FF has been shown to be well correlated with steatosis extent [11–13]. Moreover, changes in liver mechanical properties associated with hepatic diseases can be quantified by either ultrasound or MR elastographic imaging technologies [14]. Compared with ultrasound, liver MR elastography (MRE) has better accuracy and fewer technical failures, and there is an emerging consensus that hepatic MRE is the most reliable noninvasive method for detecting and staging liver fibrosis as an alternative to liver biopsy [12, 15, 16]. Despite strong co-occurrence of ballooning and fibrosis in NAFLD, liver

stiffness has also shown promising trends to correlate with ballooning grade in several preclinical and clinical investigations [12, 17–19]. Furthermore, MRE-derived loss modulus or damping ratio (relative viscous to elastic behavior) was demonstrated to be associated with inflammation in five different preclinical models of chronic liver diseases [20].

Given the aforementioned evidence, these three imaging biomarkers (FF, liver stiffness, and loss modulus) are expected to be potential predictors, either independently or jointly, of three key histological features (steatosis, ballooning, and inflammation). We expect further development of multiparametric hepatic imaging would facilitate NAFLD diagnosis and treatment monitoring by providing a noninvasive approach with accuracy as good as or better than the invasive reference standard. Based on the histologic features of NAFLD, the NAFLD activity scoring (NAS score) system is a well-accepted standard used for assessing NASH severity and measuring changes in NAFLD during therapeutic trials. According to the NASH Clinical Research Network classification, the NAS score is the sum of the steatosis grade, lobular inflammation grade, and hepatocyte ballooning grade [21]. A predictive model comprising imaging biomarkers that are sensitive to steatosis, inflammation, and ballooning may well be expected to provide a virtual NAS score (vNAS) capable of noninvasively estimating the NAS score in subjects with simple steatosis/NASH. Therefore, the purpose of this study is to evaluate the potential of MRI-assessed FF and MRE-assessed liver stiffness and loss modulus as predictors of the NAS score in a preclinical NAFLD model.

Materials and methods

Animals

With institutional animal care and use committee approval, a total of 89 C57BL/6 wild-type male mice were used in this study, including a training cohort of 64 mice and an independent testing cohort of 25 mice (Fig. 1). Based on a well-established preclinical model [20], the progressive NAFLD was developed by feeding with a fast-food diet and fructose water for 48 weeks and regressive NAFLD was developed by feeding with fast-food diet and fructose water for 24 weeks and then treated by changing to normal food and water for 24 weeks. The normal controls were fed with normal food and water. In the training cohort, the in vivo MRI/MRE data and histologic analysis of liver biopsies obtained after euthanasia were collected at 1, 12, 24, 36, and 48 weeks (Fig. 1(a)). In the testing cohort, the in vivo MRI/MRE data and histologic analysis were performed at the endpoint time after 48 weeks (Fig. 1(b)).

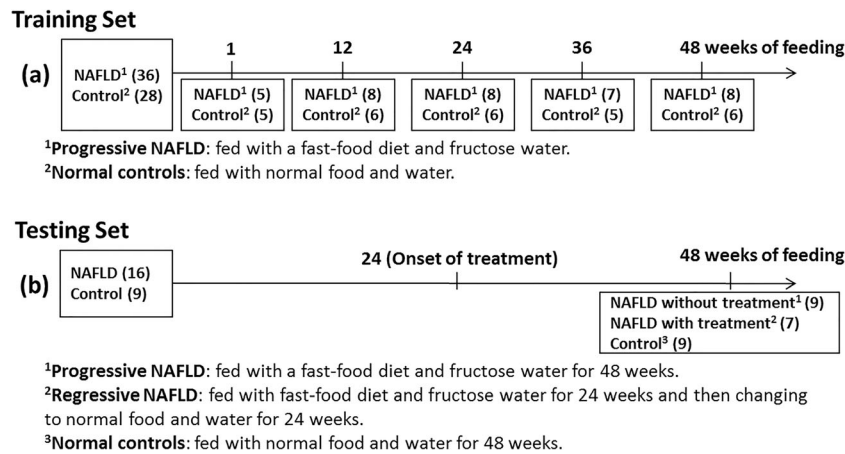


Fig. 1 Timeline of data collection. **a** Sixty-four mice in the training set had MRI/MRE exams before five specified euthanasia time points and histologic analysis of liver biopsies obtained after euthanasia. The number of mice for each group is indicated in parentheses. **b** Twenty-five mice

in the testing set had bi-weekly to monthly longitudinal MRI/MRE exams. Tissue harvesting for histologic analysis was performed only at the end point time after 48 weeks of feeding

MRI/MRE scans

The MRI/MRE scans were performed using a 3-T whole-body scanner (GE Healthcare) [22]. Briefly, after preparation and administration of maintenance anesthesia with isoflurane, each mouse was placed in a plastic cradle in the supine position, then slid into a custom eight-channel birdcage imaging coil. A disposable silver acupuncture needle with a 0.26-mm diameter and 39-mm length (Asahi Medical Instrument) was inserted into the liver through the anterior body wall. The other end of the needle was connected to a passive pneumatic driver that generated longitudinally oriented sinusoidal vibrations at 80 Hz. MRE phase (“wave”) images were acquired with a free-breathing, spin-echo echo-planar-imaging MRE sequence. Hepatic FF (%) was measured using a two-point Dixon method [23]. Details of imaging protocols are given in Appendix-E1.

Histologic analysis

Histologic analysis was performed with hematoxylin-eosin [24] and picrosirius red staining [25] of formalin-fixed, paraffin-embedded 5- μ m liver slices. Perl’s stain was performed to determine whether the iron accumulates in this model and affects the MRE measurements (see Appendix-E2). The histologic features were assessed with the NASH clinical research network scoring system [21] (T.M., a pathologist with 7 years of experience in liver histology and blinded to the MR results). The NAS score was calculated as the unweighted sum of the scores for steatosis (0–3), lobular inflammation (0–3), and ballooning (0–2). Based on the NAS score, NAFLD was classified as “not-NASH” (NAS < 3), “borderline-NASH” (NASH = 3–4), and “definite-NASH” (NAS = 5–8). Fibrosis stage was also evaluated (0, 1a, 1b,

1c, 2, 3, 4). For statistical analysis, fibrosis stages 1a, 1b, and 1c were assigned as 0.5, 1, and 1.5, respectively.

Imaging and data processing

All MRE wave data were analyzed and inverted with a direct inversion of the Helmholtz equation to calculate the complex shear modulus $G^* = G' + iG''$ (details in Appendix-E1) [20]. The shear stiffness ($|G^*|$), storage modulus (G'), loss modulus (G''), and damping ratio ($\zeta = G'' / (2 \times G')$) were calculated. Single volumetric regions of interest (ROIs) were drawn manually by two experienced readers (Z.Y., an MRI scientist with 2 years of experience in liver MRE, supervised by M.Y., an MRI scientist with > 10 years of experience in liver MRE). The ROI criteria were as follows: (a) including liver parenchyma only, (b) excluding regions without visually adequate magnitude signal or shear wave amplitude, (c) excluding the location of the vibrating needle and the adjacent area (circular area with a 3-pixel radius), and (d) staying 2 pixels away from the edges and excluding the top and bottom slices of the liver.

Statistical analysis

Power analysis was performed to determine the sample size for the training and testing groups (details in Appendix-E3). Continuous data were summarized as means and standard deviations. Categorical data were summarized as counts and percentages. Three imaging predictors were selected with Spearman’s correlation and univariate analyses (details in Appendix-E4). In the training cohort, pairwise, nonparametric comparisons with the Dunn method for joint ranking were used to compare selected imaging parameters with the severity of the components of NAS (steatosis, lobular inflammation, and ballooning). Multivariate analyses were performed to assess the contribution of imaging predictors to each individual

histologic feature. Parameter estimates and p values were reported.

In the predictive model fittings, a vNAS score was generated by fitting an ordinal logistic model (OLM) to training data with FF, G'' , and $|G^*|$ as predictor variables. The predictive accuracy of the vNAS-OLM score was estimated from the area under receiver operating characteristic (ROC) curves (AUROC) for distinguishing each NAS score. Despite the ordinal nature of the NAS score, we also separately treated the NAS score as a continuous variable; thus, a continuous vNAS score was generated by fitting a generalized linear model (GLM) to the training data.

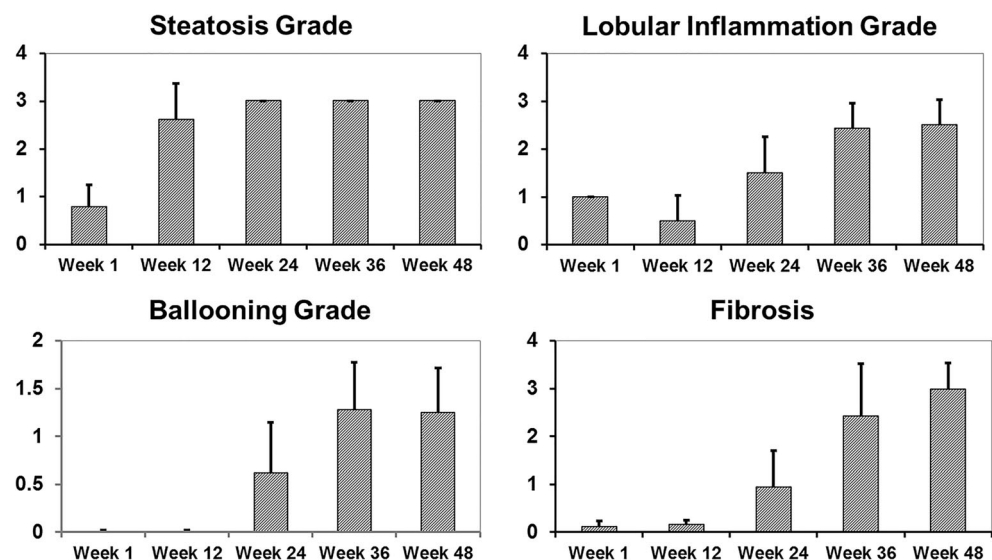
Finally, the vNAS-OLM and vNAS-GLM scores were applied to the testing cohort to validate the predictive performance of the OLM and GLM models in diagnosing NAFLD. A p value < 0.05 was considered significant. All statistical analyses were performed using JMP Pro version 12.2.0 (SAS Institute, Inc.).

Results

Histologic results

The histological changes over time in the training cohort are summarized in Fig. 2. Steatosis extent reached a peak at week 24 and remained high afterward. Lobular inflammation occurred as early as week 1, decreased at week 12, then increased at week 24 and afterward. Hepatocellular ballooning started to be present at week 24 and increased in severity afterward. There was at most minimal hepatic fibrosis before week 12, mild fibrosis at week 24, and increasingly developed moderate to severe fibrosis from weeks 36 to 48. There were minimal necroinflammation, no steatosis, and no fibrosis observed for control groups.

Fig. 2 Changes in steatosis, lobular inflammation, ballooning grades, and fibrosis stages over time in NAFLD mice in the training cohort. The values are reported as mean \pm standard deviation. The numbers of animals for each week are as follows: week 1 ($n = 5$), week 12 ($n = 8$), week 24 ($n = 8$), week 36 ($n = 7$), and week 48 ($n = 8$)



Relationships between MRI/MRE and histologic results

As shown in Fig. 3a, there was a progressive elevation in G'' with increased lobular inflammation (I0–I2) in the training cohort. There was no significant difference in G'' between the moderate and severe inflammation groups (I2 and I3). The $|G^*|$ increased significantly with the emergence of hepatocellular ballooning (Fig. 3b) and FF increased significantly with steatosis (Fig. 3c). Animals with increased steatosis also tended to have elevated $|G^*|$.

The effects of the imaging combination in predicting histologic features in the training cohort are summarized in Table 1. Both FF and $|G^*|$ had statistically significant positive effects on steatosis prediction. However, there is no evidence for an effect from G'' in predicting steatosis. The $|G^*|$ had a significant positive effect in predicting both lobular and portal inflammation, while G'' had significant positive effects in predicting lobular, but not portal inflammation. Both FF and $|G^*|$ had significant positive effects for predicting hepatocellular ballooning and subsequent fibrosis. Figure 4 shows histologic images and corresponding FF, $|G^*|$, and G'' maps in mice with different stages of NAS score.

NAS score prediction with MRI/MRE

Figure 5 shows the ROC analyses of NAS score prediction using the vNAS-OLM score in both training and testing cohorts. Each ROC curve is calculated by considering a specific NAS level or higher level (e.g., $NAS \geq 6$) to be a positive response and a lower level (e.g., $NAS < 6$) to be a negative response. The predicted vNAS-OLM score has excellent accuracy with AUROCs > 0.84 for histological $NAS > 1$ in the training and testing cohorts, except for $NAS = 5$ in the training set, and $NAS = 3–5.8$ in the testing set due to the lack of data.

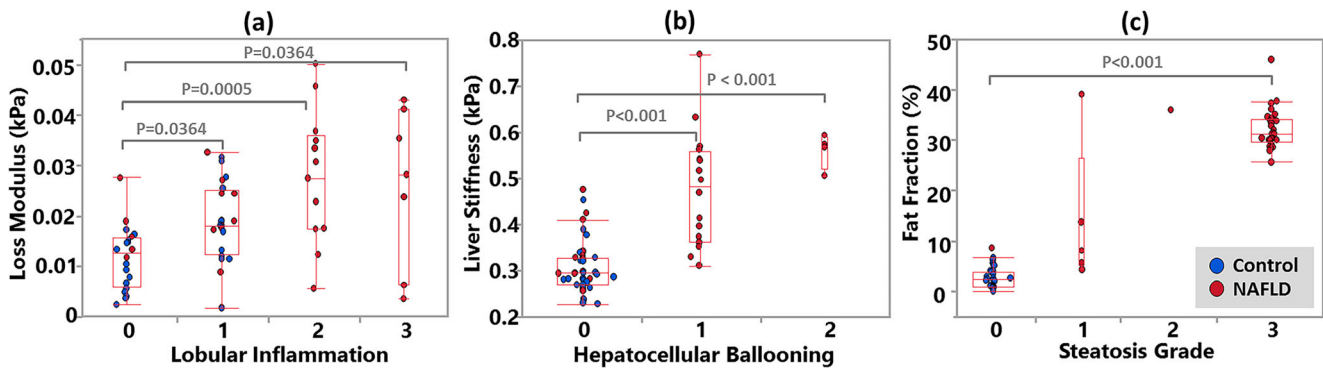


Fig. 3 Changes in MRI and MRE parameters with the severity of different histologic features in the training cohort. Scatter plots of the pairwise comparisons of three imaging parameters used to distinguish grades of three histologic features. **a** There was a progressive elevation in the loss modulus with increased lobular inflammation. **b** The liver stiffness increased significantly with the emergence of hepatocellular ballooning. **c** The fat fraction increased significantly with steatosis. Significant differences were labeled in the plots with gray square

brackets and *p* values. The control and diseased mice are illustrated in blue and red dots, respectively. The superimposed box plots indicate the 75%, median, and 25% quartiles. The interquartile range (IQR) is defined as the difference between the 75% and 25% quartiles. The extended lines (whiskers) are 75% + 1.5 × IQR and 25% − 1.5 × IQR. If the data points do not reach the computed ranges, then the whiskers are determined by the upper and lower data point values excluding outliers

The scatter plots of the predicted vNAS-OLM and vNAS-GLM scores versus the actual histologic NAS score in the training and testing cohorts are shown in Fig. 6. In the training dataset, the overall agreement between the predicted vNAS-OLM and histologic NAS score is 51/64 (80%), and the agreement for each subgroup is 33/34 (97%, not-NASH), 0/10 (0%, borderline-NASH), and 18/20 (90%, definite-NASH), respectively. The corresponding agreement between the predicted vNAS-GLM and histologic NAS score is 54/64 (84% overall), 33/34 (97% not-NASH), 6/10 (60% borderline-NASH), and 15/20 (75% definite-NASH), respectively. The training set exhibits the lowest agreement in the borderline-NASH group for both predictive models. For the testing set, the overall agreement between the histologic NAS score and predicted vNAS-OLM is 23/25 (92% overall), 17/17 (100% not-

NASH), 0/0 (no borderline-NASH presented), and 6/8 (75% definite-NASH), respectively, while for the predicted vNAS-GLM, the agreement is 23/25 (92% overall), 17/17 (100% not-NASH), 0/0 (no borderline-NASH presented), and 6/8 (75% definite-NASH), respectively.

Discussion

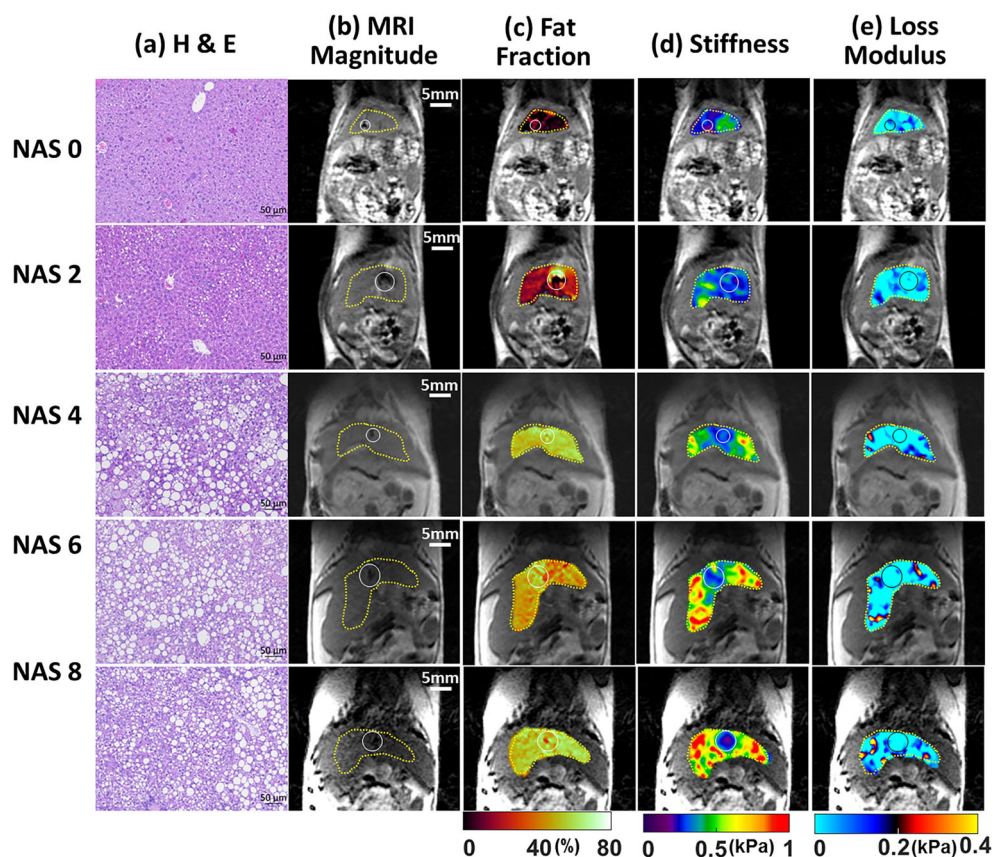
In this preclinical model, we found that the FF and loss modulus were highly related to steatosis and inflammation respectively, while liver stiffness was associated with disease severity including fibrosis, inflammation, and ballooning as expected, which agreed well with other studies [11–13, 15–18, 20]. Some of the early stage pathophysiologic processes (e.g.,

Table 1 Multivariate analyses of the three imaging parameter (fat fraction, shear stiffness, and loss modulus) effects for the prediction of several key histologic features in the training cohort of 64 mice. Effects are reported as parameter estimates, 95% confidence intervals (CI) and

p values (italicized if significant). All observed significant effects show positive correlations between the imaging predictors and the histological findings

Multivariate analysis of imaging biomarkers		Histological features in NAFLD progression				
		Steatosis	Inflammation		Ballooning	Fibrosis
			Lobular	Portal		
Fat fraction	Estimate	<i>8.08</i>	1.15	0.12	<i>0.84</i>	<i>1.55</i>
	95% CI	<i>(7.22, 8.93)</i>	(−0.13, 2.43)	(−0.46, 0.69)	<i>(0.08, 1.59)</i>	<i>(0.17, 2.93)</i>
	<i>p</i> value	<i>< 0.0001</i>	0.07	0.69	<i>0.03</i>	<i>0.03</i>
Loss modulus	Estimate	5.06	<i>19.95</i>	6.59	7.14	<i>19.76</i>
	95% CI	(−5.35, 15.47)	<i>(4.40, 35.50)</i>	(−0.39, 13.57)	(−2.01, 16.29)	<i>(2.98, 36.55)</i>
	<i>p</i> value	0.34	<i>0.01</i>	0.06	0.12	<i>0.02</i>
Shear stiffness	Estimate	2.02	4.29	<i>1.50</i>	3.00	6.35
	95% CI	(0.93, 3.11)	(2.66, 5.92)	<i>(0.77, 2.24)</i>	(2.04, 3.96)	(4.59, 8.11)
	<i>p</i> value	<i>< 0.001</i>	<i>< 0.0001</i>	<i>0.0001</i>	<i>< 0.0001</i>	<i>< 0.0001</i>

Fig. 4 Histologic analyses, MR imaging, and elastography results in representative mice with NAS scores of 0, 2, 4, 6, and 8. **a** Histologic images stained with hematoxylin-eosin (H&E). **b** MRI magnitude images and **(c)** MRI-measured fat fraction maps. MRE-measured **(d)** liver stiffness and **(e)** loss modulus maps at 80 Hz. Liver ROIs are delineated with yellow dotted lines. Locations of the vibrating needle (white circles) were excluded from the calculation

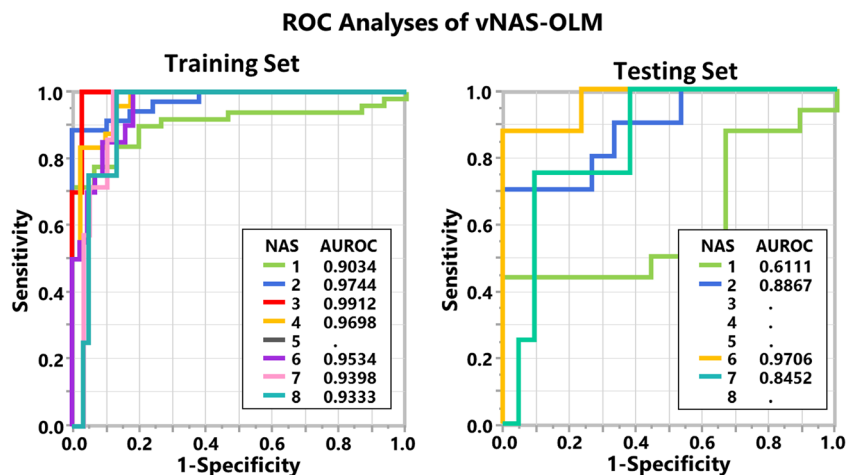


steatosis or inflammation) became progressively more severe, eventually leading to the onset of later stage disease processes (e.g., ballooning or fibrosis). Since both FF and liver stiffness increase with the disease progress, it is not surprising to observe significant correlations between these two imaging parameters and the histologic features that compose the NAS score. This finding is consistent with Imajo et al's clinical study where MRE-measured liver stiffness was significantly correlated with inflammatory and ballooning grades and MRI-measured FF was significantly correlated with steatosis and

fibrosis stages [12], though it is difficult to separate the effects of ballooning and fibrosis on liver stiffness. FF shows no significant correlation with inflammation. It has been hypothesized that a number of diverse parallel processes might contribute to the development of inflammation in NASH. Inflammation could have short-term fluctuations over the course of NAFLD progression, as opposed to steatosis and fibrosis that increasingly accumulate in this NAFLD model.

While significantly correlated, the relationships between these imaging parameters and the histologic findings may

Fig. 5 The receiver operating characteristic (ROC) analyses for NAS score prediction in the training (left) and testing (right) cohorts. An ordinal logistical model was used to predict NAS scores (vNAS-OLM) for ROC analysis. The areas under the ROC curve (AUROC) are listed in the table in the lower right



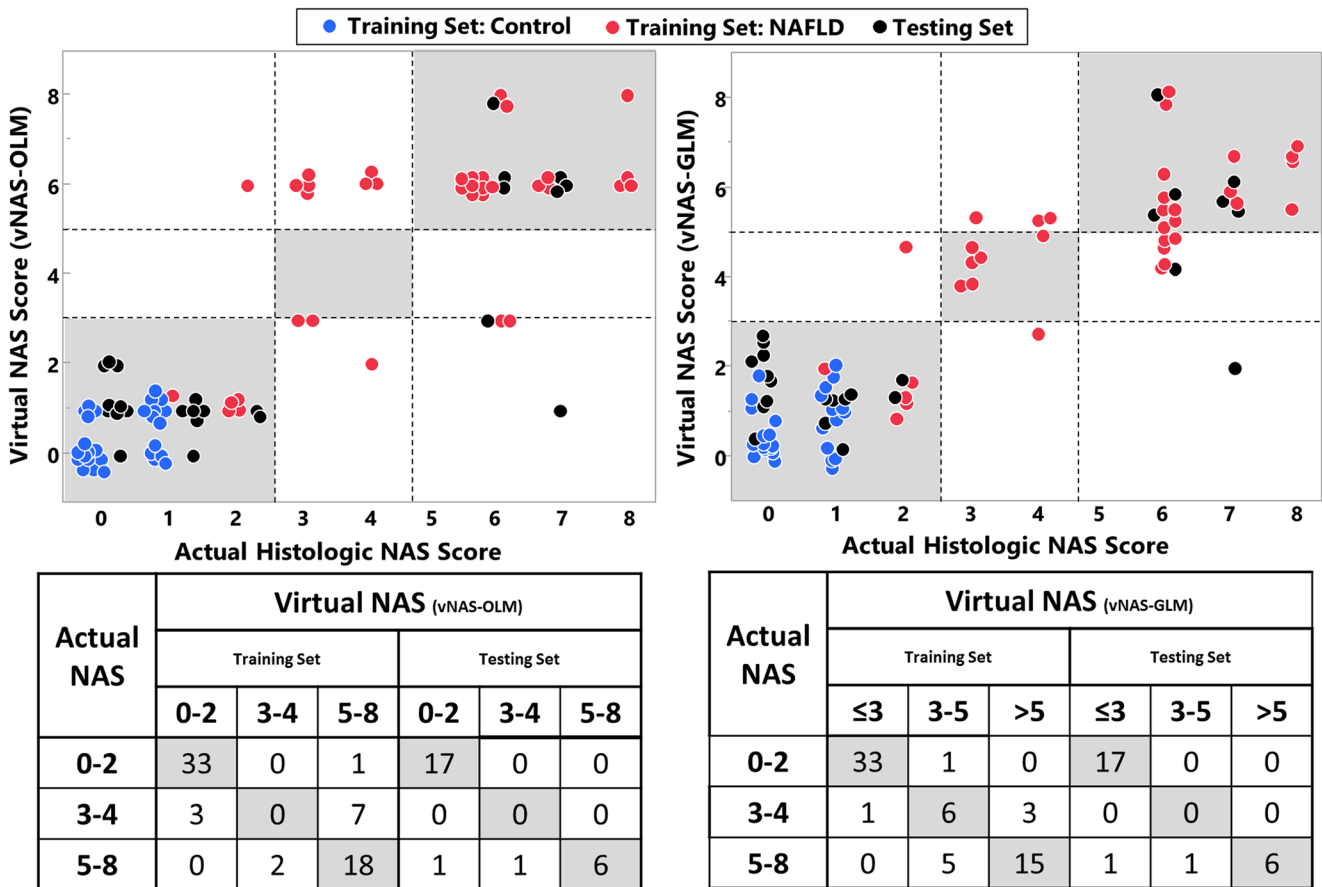


Fig. 6 Scatter plots of the predicted vNAS-OLM (top left) and vNAS-GLM (top right) scores and the actual histologic NAS score in the training cohort of 64 mice (blue and red dots) and the testing cohort of 25 mice (black dots) with the corresponding 3 × 3 agreement tables shown below. This vNAS-GLM score was calculated as (vNAS-GLM = 10.1 × FF +

32.2 × G'' + 9.3 × | G^* | − 2.8; FF—fat fraction, G'' —loss modulus, | G^* |—liver stiffness). Note that the vNAS-GLM scores are continuous, rather than discrete values. For both predictive models, the points are slightly shifted along the horizontal coordinate for visibility

not be as simple as monotonic linear functions. In practice, these pathophysiologic events can coexist and interactively affect each other during the NAFLD progression or regression. Medical diagnostic decisions are often based on a histological scoring system such as the widely used NAS score that quantifies steatosis, ballooning, and inflammation, or a more recently introduced SAF score that encompasses an assessment of steatosis (*S*), activity (*A*, the unweighted sum of ballooning and lobular inflammation), and fibrosis (*F*) [26]. Although many clinical investigations have shown that the subjective NAS score has only fair to moderate repeatability and reproducibility [21, 27–29], and that the SAF decision tree diagnosis may be more accurate in identifying NASH, the NAS score tends to be more suitable to fit within the clinical trials for assessing changes in NAFLD severity given its large dynamic range from 0 to 8 to evaluate severity of liver injury. Therefore, this study has selected the NAS score as the reference standard to develop the multiparametric imaging model.

Excellent overall agreement and accuracy of both vNAS-OLM and vNAS-GLM prediction to the actual histological

NAS score were observed in the training dataset. However, the training set had the lowest agreement for the borderline-NASH group with both predictive models, which is not surprising as there exists a wide gray zone (NAS 3–4) where NASH may or may not be present, and the use of NAS score to diagnosis borderline-NASH remains controversial [21, 30]. Due to the lack of the borderline-NASH group in the testing dataset, the model validation only focused on the performance for the “not-NASH” and “definite-NASH” groups. The training dataset has an overall agreement of 80% and 84% for these two groups by using OLM and GLM prediction respectively, while agreement for the testing dataset is 92% with both OLM and GLM prediction, respectively. For the “not-NASH” and “definite-NASH” diagnoses, the testing dataset successfully validated the performance of the proposed imaging predictive models. Our preliminary results showed that continuous vNAS-GLM had slightly better predictive performance compared with the categorical vNAS-OLM score, as it provides better accuracy and precision with the expected tendency to follow the actual progressive disease development. The misclassifications may be because (1) the histologically assessed

NAS score is subjective and has potential sampling errors and (2) some controls have mild lobular inflammation, which may be due to the invasive MRE needle penetration method adopted in this study.

In this study, the NAFLD mouse model has a different disease progression compared with NAFLD patients. For example, many NAFLD mice have steatosis as early as week 1, whereas NASH patients may present with little to no steatosis, suggesting that inflammation may occur first [31]. In this situation, inflammation results in a stress response of the hepatocytes, which may lead to lipid accumulation, and therefore could precede steatosis in NASH. Hepatic steatosis may be considered as a bystander phenomenon subsequent to inflammatory attacks. On the other hand, NASH subsequent to simple steatosis may be the consequence of persistent and promoted inflammation. Many of these pathophysiologic events may take place in parallel rather than consecutively as in the mouse model, therefore not allowing the exact determination of individual trends and effects in the evolution of NAFLD. However, the study of *in vivo* animal models has many advantages. First, it provides the well-controlled etiologic background and extrahepatic effects, the ability to allow examination of pathophysiological status, and detailed histological analysis, which are all unavailable in human studies. Second, human subjects usually have a long chronic disease progression, which is impractical to follow in the study design. Moreover, recruited human subjects are often in a chronic stage of the disease, making it difficult to measure the effects of disease onset. Therefore, the use and outcome of this animal model are pivotal to bridge the translational gap to the clinic.

This study still has several limitations. First, this study is limited to a single diet-induced NAFLD model. In practice, NAFLD frequently coexists with other liver diseases (e.g., viral hepatitis, hemochromatosis, and alcoholic liver disease). Thus, more studies involving different etiologies and disease states are needed to further validate the relationships between imaging predictors and histologic findings before clinical translation. Second, given that no iron overload was observed in this preclinical model (Fig. A), the FF was calculated based on a gradient-echo sequence with two echoes. Improvements in pulse sequence programming, including T1, T2*, and multi-peak fatty component correction, will be required in our future studies. Third, the use of NAS score to diagnosis borderline-NASH has remained controversial [21, 30]. The capabilities of our imaging biomarkers to predict other scoring systems (e.g., SAF score [26]) for NASH diagnosis need to be further investigated. Fourth, we did not achieve a sufficient sample size for steatosis grades 1 and 2 and ballooning grade 2 in this training cohort and NAS scores 3–5 in the testing cohort, which limits the value of the validation, especially for the borderline-NASH prediction. Finally, although the linear model performed well in this preclinical model, future studies

incorporating data from other preclinical and clinical models and more sophisticated statistical techniques (e.g., nonlinear machine learning algorithms) will be performed to further optimize and validate the predictive values of our vNAS imaging indicators. Many studies have demonstrated high repeatability and reproducibility of MRI/MRE measurements [32–37]. Although different acquisition strategies would lead to different regression coefficients, the concept of the imaging-based vNAS score is highly translational to other MR studies. Therefore, we expect that the optimized, imaging-based vNAS score would provide higher reliability in noninvasively assessing liver disease severity and treatment efficacy.

Conclusion

This is the first study in a NAFLD preclinical model showing that imaging biomarkers from multiparametric hepatic MRI/MRE can provide an accurate, noninvasive prediction of the NAFLD activity score. The proposed model could be rapidly applied to other preclinical and clinical studies to cross-validate the relationships systematically and may have important applications in drug evaluation, disease monitoring, and therapy response assessment.

Funding This work has been supported by National Institutes of Health (NIH) grants EB017197, EB001981, and DK111378.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Richard L. Ehman, M.D., Department of Radiology, Mayo Clinic.

Conflict of interest The Mayo Clinic and the authors of this manuscript have intellectual property and a financial interest related to this research.

This research has been reviewed by the Mayo Clinic Conflict of Interest Review Board and is being conducted in compliance with the Mayo Clinic Conflict of Interest policies.

Statistics and biometry Two of the authors (Terry M. Therneau and Heshan Liu) are senior statisticians.

Informed consent Written informed consent was not required for this study because this is an animal study.

Ethical approval Institutional Review Board approval was obtained. Approval from the institutional animal care committee was obtained.

Study subjects or cohorts overlap Some study subjects or cohorts have not been previously reported.

Methodology

- prospective
- experimental
- performed at one institution

References

- Browning JD, Szczepaniak LS, Dobbins R et al (2004) Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology* 40:1387–1395
- Younossi ZM, Blissett D, Blissett R et al (2016) The economic and clinical burden of nonalcoholic fatty liver disease in the United States and Europe. *Hepatology* 64:1577–1586
- Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M (2016) Global epidemiology of nonalcoholic fatty liver disease—meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 64:73–84
- Ehman RL (2009) Science to practice: can MR elastography be used to detect early steatohepatitis in fatty liver disease? *Radiology* 253:1–3
- Charlton M (2004) Nonalcoholic fatty liver disease: a review of current understanding and future impact. *Clin Gastroenterol Hepatol* 2:1048–1058
- Adams LA, Lindor KD, Angulo P (2004) The prevalence of auto-antibodies and autoimmune hepatitis in patients with nonalcoholic fatty liver disease. *Am J Gastroenterol* 99:1316–1320
- Schwenzer NF, Springer F, Schraml C, Stefan N, Machann J, Schick F (2009) Non-invasive assessment and quantification of liver steatosis by ultrasound, computed tomography and magnetic resonance. *J Hepatol* 51:433–445
- Cassidy FH, Yokoo T, Aganovic L et al (2009) Fatty liver disease: MR imaging techniques for the detection and quantification of liver steatosis. *Radiographics* 29:231–260
- Reeder SB, Sirlin CB (2010) Quantification of liver fat with magnetic resonance imaging. *Magn Reson Imaging Clin N Am* 18:337–357 ix
- Kramer H, Pickhardt PJ, Kliever MA et al (2017) Accuracy of liver fat quantification with advanced CT, MRI, and ultrasound techniques: prospective comparison with MR spectroscopy. *Am J Roentgenol* 208:92–100
- Tang A, Desai A, Hamilton G et al (2015) Accuracy of MR imaging-estimated proton density fat fraction for classification of dichotomized histologic steatosis grades in nonalcoholic fatty liver disease. *Radiology* 274:416–425
- Imajo K, Kessoku T, Honda Y et al (2016) Magnetic resonance imaging more accurately classifies steatosis and fibrosis in patients with nonalcoholic fatty liver disease than transient elastography. *Gastroenterology* 150(626–637):e627
- Paige JS, Bernstein GS, Heba E et al (2017) A pilot comparative study of quantitative ultrasound, conventional ultrasound, and MRI for predicting histology-determined steatosis grade in adult nonalcoholic fatty liver disease. *AJR Am J Roentgenol* 208:W168–W177
- Chen J, Yin M, Talwalkar JA et al (2017) Diagnostic performance of MR elastography and vibration-controlled transient elastography in the detection of hepatic fibrosis in patients with severe to morbid obesity. *Radiology* 283:418–428
- Yin M, Glaser KJ, Talwalkar JA, Chen J, Manduca A, Ehman RL (2016) Hepatic MR elastography: clinical performance in a series of 1377 consecutive examinations. *Radiology* 278:114–124
- Song P, Mellema DC, Sheedy SP et al (2016) Performance of 2-dimensional ultrasound shear wave elastography in liver fibrosis detection using magnetic resonance elastography as the reference standard: a pilot study. *J Ultrasound Med* 35:401–412
- Rausch V, Peccerella T, Lackner C et al (2016) Primary liver injury and delayed resolution of liver stiffness after alcohol detoxification in heavy drinkers with the PNPLA3 variant I148M. *World J Hepatol* 8:1547–1556
- Ogawa S, Moriyasu F, Yoshida K et al (2016) Relationship between liver tissue stiffness and histopathological findings analyzed by shear wave elastography and compression testing in rats with non-alcoholic steatohepatitis. *J Med Ultrason* 43:355–360
- El Bokl MA-EH, Mabrouk SH, El Karmoty KZ, Bakir ASM, El Aleem AMA (2012) Fibroscan as a noninvasive tool in the assessment of the degree of hepatic steatosis in patients with nonalcoholic fatty liver disease. *Egypt Liver J* 2:83–87
- Yin M, Glaser KJ, Manduca A et al (2017) Distinguishing between hepatic inflammation and fibrosis with MR elastography. *Radiology* 284:694–705
- Kleiner DE, Brunt EM, Van Natta M et al (2005) Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 41:1313–1321
- Yin M, Woollard J, Wang X et al (2007) Quantitative assessment of hepatic fibrosis in an animal model with magnetic resonance elastography. *Magn Reson Med* 58:346–353
- Hayashi T, Saitoh S, Takahashi J et al (2017) Hepatic fat quantification using the two-point Dixon method and fat color maps based on non-alcoholic fatty liver disease activity score. *Hepatol Res* 47:455–464
- Fischer AH, Jacobson KA, Rose J, Zeller R (2008) Hematoxylin and eosin staining of tissue and cell sections. *CSH Protoc* 2008:prot4986
- Junqueira LC, Bignolas G, Brentani RR (1979) Picrosirius staining plus polarization microscopy, a specific method for collagen detection in tissue sections. *Histochem J* 11:447–455
- Bedossa P, Poitou C, Veyrie N et al (2012) Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology* 56:1751–1759
- Juluri R, Vuppalanchi R, Olson J et al (2011) Generalizability of the nonalcoholic steatohepatitis Clinical Research Network histologic scoring system for nonalcoholic fatty liver disease. *J Clin Gastroenterol* 45:55–58
- Poumik O, Alavian SM, Ghalichi L et al (2014) Inter-observer and intra-observer agreement in pathological evaluation of non-alcoholic fatty liver disease suspected liver biopsies. *Hepat Mon* 14:e15167
- Sumida Y, Nakajima A, Itoh Y (2014) Limitations of liver biopsy and non-invasive diagnostic tests for the diagnosis of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *World J Gastroenterol* 20:475–485
- Brunt EM, Kleiner DE, Wilson LA, Belt P, Neuschwander-Tetri BA (2011) Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings. *Hepatology* 53:810–820
- Tiniakos DG, Vos MB, Brunt EM (2010) Nonalcoholic fatty liver disease: pathology and pathogenesis. *Annu Rev Pathol* 5:145–171
- Kang GH, Cruite I, Shieh-morteza M et al (2011) Reproducibility of MRI-determined proton density fat fraction across two different MR scanner platforms. *J Magn Reson Imaging* 34:928–934
- Middleton MS, Haufe W, Hooker J et al (2017) 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* 283:438–449
- Lee Y, Lee JM, Lee JE et al (2014) MR elastography for noninvasive assessment of hepatic fibrosis: reproducibility of the examination and reproducibility and repeatability of the liver stiffness value measurement. *J Magn Reson Imaging* 39:326–331
- Yasar TK, Wagner M, Bane O et al (2016) Interplatform reproducibility of liver and spleen stiffness measured with MR elastography. *J Magn Reson Imaging* 43:1064–1072
- Shire NJ, Yin M, Chen J et al (2011) Test-retest repeatability of MR elastography for noninvasive liver fibrosis assessment in hepatitis C. *J Magn Reson Imaging* 34:947–955
- Serai SD, Obuchowski NA, Ventakesh SK et al (2017) Repeatability of MR elastography of liver: a meta-analysis. *Radiology* 285:92–100

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.