Title Page

A Novel Quantitative Ultrasound Technique for Identifying Nonalcoholic Steatohepatitis

Short Title: Ultrasound score for NASH

Authors' name:

Feng Gao, MD[1]; Qiong He, PhD[2,3]; Gang Li, MD[4]; Ou-Yang Huang, MD[4]; Liang-Jie Tang, MD[4]; Xiao-Dong Wang[5]; Giovanni Targher, MD[6]; Christopher D. Byrne, MD, PhD [7]; Jian-Wen Luo, PhD[2]; Ming-Hua Zheng, MD, PhD [4,5,8*]

Affiliations:

[1]Department of Gastroenterology, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

[2]Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China

[3]Tsinghua-Peking Joint Center for Life Sciences Department, Tsinghua University, Beijing 100084, China

[4]NAFLD Research Center, Department of Hepatology, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

[5]Key Laboratory of Diagnosis and Treatment for The Development of Chronic Liver Disease in Zhejiang Province, Wenzhou, China

[6]Section of Endocrinology, Diabetes and Metabolism, Department of Medicine, University and Azienda Ospedaliera Universitaria Integrata of Verona, Verona, Italy

[7]Southampton National Institute for Health Research Biomedical Research Centre,

University Hospital Southampton, Southampton General Hospital, Southampton, UK

[8]Institute of Hepatology, Wenzhou Medical University, Wenzhou, China

***Corresponding authors:**

Ming-Hua Zheng, MD, PhD

NAFLD Research Center, Department of Hepatology, the First Affiliated Hospital of Wenzhou Medical University; No. 2 Fuxue Lane, Wenzhou 325000, China.

E-mail: zhengmh@wmu.edu.cn; fax: (86) 577-55578522; tel: (86) 577-55579622.

Abstract word count: 220 words

Total word count: 2801 words (main text, *excluding* abstract, references and figure legends)

Number of figures: 4

Number of tables: 4

Abbreviations

ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; UAP, ultrasound attenuation parameter; CI, confidence interval; CK-18, cytokeratin-18 fragments; GGT, γ-glutamyltranspeptidase; HOMA-IR, homeostasis model assessment of insulin resistance; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; LSM, liver stiffness measurement; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; NAS-CRN, NASH-Clinical Research Network; NAS, NAFLD activity score; NFS, NAFLD fibrosis score; NPV, negative predictive value; PPV, positive predictive value; QUS,

quantitative ultrasound; TC, total cholesterol; TE, transient elastography; TG, triglycerides.

Article guarantor: Ming-Hua Zheng, MD, PhD

Authors' Contributions

Study concept and design: Feng Gao and Ming-Hua Zheng

Acquisition of data: Qiong He; Gang Li; Ou-Yang Huang; Liang-Jie Tang; Xiao-Dong Wang

Drafting of the manuscript: Feng Gao and Qiong He

Critical revision: Giovanni Targher and Christopher D. Byrne

Statistical analysis: Feng Gao and Qiong He

Study supervision: Ming-Hua Zheng and Jian-Wen Luo

All authors contributed to the manuscript for important intellectual content and approved the final version of the manuscript.

Conflict of interest statement:

All authors: nothing to declare.

Acknowledgments

We thank Herui Biomed Company Limited (Suzhou, China) for providing CK-18 M30 ELISA kits.

Financial support

This work was supported by grants from the National Natural Science Foundation of China (82070588), High Level Creative Talents from Department of Public Health in

Zhejiang Province (S2032102600032) and Project of New Century 551 Talent Nurturing in Wenzhou. GT is supported in part by grants from the School of Medicine, University of Verona, Verona, Italy. CDB is supported in part by the Southampton NIHR Biomedical Research Centre (IS-BRC-20004), QH and JL were supported in part by the National Natural Science Foundation of China (61801261), Tsinghua-Peking Joint Center for Life Sciences and the Young Elite Scientists Sponsorship by China Association for Science and Technology.

ABSTRACT

BACKGROUND & AIMS: There remains a need to develop a non-invasive, accurate and easy-to-use tool to identify patients with nonalcoholic steatohepatitis (NASH). Successful clinical and preclinical applications demonstrate the ability of quantitative ultrasound (QUS) techniques to improve medical diagnostics. We aimed to develop and validate a diagnostic tool, based on QUS analysis, for identifying NASH.

METHODS: A total of 259 Chinese individuals with biopsy-proven non-alcoholic fatty liver disease (NAFLD) were enrolled in the study. The histological spectrum of NAFLD was classified according to the NASH clinical research network scoring system. Radiofrequency (RF) data, raw data of *iLivTouch*, was acquired for further QUS analysis. The least absolute shrinkage and selection operator (LASSO) method was used to select the most useful predictive features.

RESULTS: Eighteen candidate RF parameters were reduced to two significant parameters by shrinking the regression coefficients with the LASSO method. We built a novel QUS-score based on these two parameters, and this QUS-score showed good discriminatory capacity and calibration for identifying NASH both in the training set (area under the ROC curve [AUROC]: 0.798, 95% CI 0.731-0.865; Hosmer-Lemeshow test, p=0.755) and in the validation set (AUROC: 0.816, 95 %CI 0.725-0.906; Hosmer-Lemeshow test, p=0.397). Subgroup analysis showed that the QUS-score performed well in different subgroups.

CONCLUSIONS: The QUS-score, which was developed from quantitative

ultrasound, provides a novel, non-invasive and practical way for identifying NASH.

Keywords: NAFLD, NASH, quantitative ultrasound, transient elastography, radiofrequency.

Introduction

Non-alcoholic fatty liver disease (NAFLD) is an epidemic and public health threat that affects ~25% of the adult population globally.[1] NAFLD is classified histologically as simple steatosis (NAFL), non-alcoholic steatohepatitis (NASH) with varying levels of fibrosis, and cirrhosis.[2] In recent years, NASH has become the most rapidly increasing indication for liver transplantation in the United States.[3] In addition, NASH is closely associated with an increased risk of developing important extra-hepatic complications, such as type 2 diabetes, cardiovascular disease, chronic kidney disease and certain extra-hepatic malignancies.[4-7] Thus, the correct identification of patients with NASH is clinically relevant for prognosis and therapy decisions.

NASH remains difficult to diagnose despite considerable research in this area in recent years.[8] Currently, there are no simple, non-invasive tests available for diagnosing NASH, specifically in the absence of, or presence of mild liver fibrosis. The current expensive 'gold standard' for diagnosing and staging NASH, both in clinical practice and in investigational trials, is liver biopsy, which is an invasive procedure with associated risks, such as post-procedural bleeding.[9] Therefore, there remains a need to develop a non-invasive, accurate, and easy-to-use tool to identify patients with NASH.

Conventional medical imaging technologies, including ultrasonography, have

continued to improve over the years. The value of quantitative ultrasound (QUS) analysis has been widely recognized, particularly for the diagnosis of liver fibrosis. Vibration-controlled transient elastography, point shear-wave elastography (pSWE), two-dimensional shear-wave elastography (2D-SWE) and acoustic radiation force impulse (ARFI) elastography are current ultrasound elastography techniques that enable the assessment of liver stiffness in real time through a quantitative electrogram.[10] These imaging techniques are a significant breakthrough for the diagnosis and staging of liver fibrosis,[11,12] but are of limited value for the diagnosis of NASH in the absence of liver fibrosis.

iLivTouch is based on transient elastography (TE), and has been widely used in the clinic, because it enables fast and quantitative assessment of liver stiffness and steatosis.[13] During the assessment with *iLivtouch*, both the ultrasound attenuation parameter and liver stiffness measurement can be easily obtained. These two parameters have proved to be valuable for measurement of liver fibrosis and steatosis.[14-16] In addition, radiofrequency (RF) raw data can be also acquired from *iLivTouch* for further QUS analysis. RF data can provide specific numbers related to tissue features that can increase the specificity of image findings leading to improvements in diagnostic ultrasound. A large number of preclinical and clinical studies have proved that QUS parameters are related to the microstructure changes within tissues.[17-23]

Therefore, the aim of our study was to develop and test the diagnostic performance of a non-invasive tool based on QUS technology for identifying the presence of NASH in a cohort of Chinese patients with biopsy-confirmed NAFLD.

Methods

Study population and design

This is a retrospective analysis of our well-characterized Prospective Epidemic Research Specifically of NASH (PERSONS) cohort.[24] For the present analysis, we initially included 279 adult individuals with suspected NAFLD, who accepted to undergo *iLivTouch* examinations between December 2016 to September 2019 at the First Affiliated Hospital of Wenzhou Medical University (China). Suspected NAFLD was defined as either the presence of imaging-defined hepatic steatosis, or persistently elevated serum transaminases or abnormal liver fibrosis tests (i.e., elevated NAFLD fibrosis score or FIB-4 index) in subjects with metabolic risk factors, such as overweight/obesity, type 2 diabetes or metabolic syndrome. [25] Subsequently, subjects with at least one of the following criteria were excluded from analyses: (i) significant alcohol consumption (≥140 g/week in men or ≥70 g/week in women, respectively); (ii) presence of viral hepatitis, autoimmune hepatitis, drug-induced liver injury or other known chronic liver diseases; (iii) unsuccessful *iLivTouch* measurements; and (iv) incomplete important clinical variables or inadequate portal tracts on liver histology. As a consequence of these exclusion criteria, a total of 259 individuals with NAFLD were included the final analysis (as detailed in **Supplementary Figure 1**).

The study protocol was approved by the ethics committee of the First Affiliated Hospital of Wenzhou Medical University and registered in the Chinese Clinical Trial Registry (ChiCTR-EOC-17013562). All procedures involving the study participants were performed in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration. All participants signed a written informed consent to participate in this study.

Clinical and biochemical data

Clinical and biochemical data were obtained from all participants within 48 hours from liver biopsy examinations. Blood samples were taken in fasting conditions. Methodological details for measurement of plasma cytokeratin-18 fragments (CK-18 neoepitope M30) concentrations have been reported previously.[26] Body mass index (BMI) was calculated using the formula weight (kilograms) divided by height (meters) squared. Obesity was defined as BMI ≥25 kg/m^2 and severe obesity was defined as BMI ≥30 kg/m^2 for Asian individuals.[27] Insulin resistance was estimated using the homoeostasis model assessment of insulin resistance (HOMA-IR) score and defined as HOMA-IR >2.5. Diabetes mellitus was diagnosed as either self-reported history of disease, a fasting glucose level ≥ 7.0 mmol/L (≥126 mg/dL), hemoglobin A1c ≥ 6.5% (≥ 48 mmol/mol) or use of any anti-hyperglycemic drugs. Hypertension was defined as blood pressure ≥140/90 mmHg or the use of any anti-hypertensive agents. Metabolic syndrome (MetS) was defined as having at least three of the

following metabolic risk factors: central obesity (waist circumference ≥85 cm in men and ≥ 80 cm in women), increased blood pressure (systolic blood pressure ≥130 mmHg or diastolic blood pressure ≥85 mmHg or use of any antihypertensive drugs), increased fasting glucose (≥5.6 mmol/L [≥100 mg/dL] or use of any anti-hyperglycemic agents), high triglycerides (>1.7 mmol/L [≥150 mg/dL] or use of any lipid-lowering drugs) and low high-density lipoprotein (HDL) cholesterol levels (<1.03 mmol/L [<40 mg/dL] in men and <1.29 mmol/L [<50 mg/dL] in women, or use of any lipid-lowering drugs).[28]

iLivTouch parameter measurements

The ultrasound attenuation parameter (UAP) and liver stiffness measurement (LSM) were measured non-invasively using the *iLivTouch* (Wuxi Hisky Medical Technologies Co., Ltd., China) by a single trained technician, who was blinded to clinical, biochemical and histological data of participants. The specific operation method was as follows: the probe was placed below the right seventh, eighth or ninth ribs in the space between the anterior and mid-axillary lines to assess the liver continuously. Ten successful reads were required and the median measurement was recorded. The ratio of the interquartile range (IQR) divided by median (IQR/median) of all measurements less than 30% with a success rate (successful tests/total tests) ≥60% was regarded as a valid measurement.

QUS techniques involve spectral-based parameterization and envelope statistics.[19] In

particular, spectral-based parameterization techniques include the estimation of the backscatter coefficient, estimation of attenuation, and estimation of scatterer properties, such as the correlation length associated with an effective scatterer diameter and the effective acoustic concentration of scatterers. Envelope statistics include the estimation of the number density of scatterers and quantification of coherent to incoherent signals produced from the tissue. Oelze et al. have provided more detailed physical and technical explanations about QUS techniques in their review.[19] According to measurement results and RF data acquired from *iLivTouch*, a total of 18 ultrasonographic features can be obtained with different time gain compensation and named as P1 to P18 based on the intensity, attenuation and scattering characteristics of the ultrasound signals, including UAP (P1) and LSM (P3) (**Fig. 1**). P2-P10 are mainly related to intensity, whereas P11-P18 are mainly related to scattering. All these parameters are frequency and attenuation dependent. The toolbox of *iLivTouch*, *LivQ*-box, was utilized to obtain these features. The *LivQ*-box can be provided by contacting the corresponding author. And the technology is applying for a patented (Patent pending No. CN110477954A).

Liver histology

An ultrasound-guided liver biopsy was performed under sedation using a 16-gauge Hepafix needle (Gallini, Modena, Italy). The minimal number of portal tracts considered sufficient for an adequate assessment of liver histology specimens was six, according to the 2014 RCPath 'Tissue Pathways' guideline.[29] The assessment of liver

histology was undertaken by one experienced histopathologist (who was blinded to the clinical and laboratory data of participants), according to the NASH-Clinical Research Network (CRN) Scoring System.[30] NAFLD activity score (NAS) was calculated as the sum of three histological components, including liver steatosis (0-3), ballooning (0-2), and lobular inflammation (0-3). Liver fibrosis was staged as zero to 4 according to the Brunt's histologic criteria. Definite NASH was defined as the presence of hepatic steatosis, lobular inflammation and ballooning with NAS ≥5.[30,31]

Statistical analysis

Continuous variables were expressed as means ± SD or medians with interquartile ranges, and compared using either the unpaired Student's *t*-test or the Mann-Whitney U test as appropriate. Categorical variables were expressed as numbers (or percentages) and compared using the chi-squared test or the Fisher's exact test as appropriate.

For the development of our novel diagnostic tool, the study population was randomly assigned in a 2:1 ratio to training and validation sets, using a split-sample method by an experienced statistician. The least absolute shrinkage and selection operator (LASSO) method, which is suitable for the regression of high-dimensional data, was used to select the most useful predictive features from the training set. A QUS-score was calculated for each participant via a linear combination of selected ultrasonographic features that were weighted by their respective coefficients. The

accuracy of the QUS-score was subsequently evaluated both in the training and validation sets by assessing its discriminatory power and calibration capability. The diagnostic discriminatory capability of the QUS-score was evaluated by calculating the area under the receiver operating characteristic curve (AUROC), and the calibration capability was assessed by the calibration curve and the Hosmer-Lemeshow goodness of fit test. The overall sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) were calculated for each cut-off of QUS-score.

Of liver texture features, 18 predictive RF features were reduced to two significant parameters by shrinking the regression coefficients with the LASSO method on the basis of the 173 patients with NAFLD included in the training set (**Fig. 2A and 2B**). These were features with nonzero coefficients in the LASSO logistic regression model. These two features were included in the following QUS-score = 0.063*P7 + 0.019*P13. In the raw RF data, P7 is mainly related to intensity, and P13 is mainly related to scattering.

Statistical analyses were two-sided and significance was set at $p<0.05$. All statistical tests were performed using R (Version 3.3.1 The R Foundation).

Results

Baseline characteristics of participants

A total of 259 individuals with biopsy-proven NAFLD were enrolled in the final analysis. The mean age of participants was 42.8 years and 72.6% of them were men. Their mean values of BMI and waist circumference were 26.7 kg/m^2 and 92.2 cm, respectively. 98 subjects (37.8%) had established type 2 diabetes, 77 (29.7%) had hypertension (29.7%), and 193 (74.5%) subjects had MetS. Their median value of the NAS score was 4 (IQR 3-5). NASH was histologically diagnosed in 100 subjects (38.6%) and NASH with any stage of fibrosis was diagnosed in 90 subjects (34.7%). By study design, 173 subjects and 86 subjects were randomly assigned to the training and validation sets, respectively. Baseline characteristics of the study participants are summarized in **Table 1**.

Development and validation of a diagnostic tool for diagnosis of NASH

The QUS-score yielded AUROCs of 0.798 (95% CI 0.731-0.865) in the training set and 0.816 (95% CI 0.725-0.906) in the validation set, which were better than those of plasma CK-18-M30 levels (AUROC 0.697, 95% CI 0.613-0.782 in the training set; and 0.753, 95%CI 0.645-0.840 in the validation set) for identifying those with NASH on liver histology (**Figs. 3A and 3B**). The calibration curves of the QUS-score for the probability of having definite NASH showed good agreement between prediction and observation both in the training and validation sets (**Figs. 3C and 3D)**. The Hosmer-Lemeshow test yielded non-significant statistics both in the training set (p = 0.755) and in the validation set (p = 0.397), which suggests that there was little departure from perfect fit.

Next, we also tested and compared the diagnostic performances of the QUS-score and plasma CK-18-M30 levels in identifying either hepatic steatosis or NASH with any histologic stage of liver fibrosis. As shown in **Table 2**, the discriminatory capability of the QUS-score was also superior to plasma CK-18-M30 levels for diagnosing steatosis, or NASH with fibrosis, both in the training and validation sets. In the training set, the AUROCs of the QUS-score was 0.907 (95 %CI, 0.860-0.954) and 0.773 (0.701- 0.845) for diagnosing steatosis and NASH with fibrosis, respectively. In the validation set, the AUROCs of the QUS-score was 0.886 (0.801- 0.971) and 0.796 (0.703- 0.889) for diagnosing steatosis and NASH with fibrosis, respectively. In addition, we found that the QUS-score increased significantly across histological grades of hepatic steatosis, ballooning, lobular inflammation, or NAS score (**Fig. 4**).

With the specific aim of identifying (by ROC curve) the most accurate cut-off values of QUS-score for ruling out and ruling in NASH, we identified two cut-off values of ≤6 and ≥7.5, respectively. In the training set, the cut-off value of QUS-score ≤6 had a sensitivity of 90.5% and a NPV of 91% to rule out NASH, whereas the cut-off value ≥ 7.5 had a specificity of 94.5% and a PPV of 75% to rule in NASH. In the validation set, the cut-off value of QUS-score ≤6 had a sensitivity of 86.5% and a NPV of 86.8% to rule out NASH, whereas the cut-off value ≥7.5 had a specificity of 89.8% and a PPV of 73.7% to rule in NASH (**Table 3**).

Subgroup analyses

We tested the diagnostic performance of the QUS-score in participants stratified by age, sex, T2DM, MetS, obesity or elevated serum ALT levels. As shown in **Table 4**, there was no significant difference in AUROC between men and women. The QUS-score performed well in patients with and without pre-existing T2DM as well as in those with and without MetS. Stratifying by age groups, the AUROC of the QUS-score for patients <60 years was 0.79 (0.73-0.85), and the AUROC for patients ≥60 years was 0.86 (0.69-1.00). Stratifying by BMI, the AUROC of the score for patients without obesity was 0.82 (0.72-0.92), the AUROC for patients with obesity was 0.82 (0.75-0.89), and the AUROC for patients with severe obesity was 0.58 (0.37-0.79). Moreover, the AUROC of the QUS-score for patients with normal ALT levels was 0.79 (0.67-0.90), and the AUROC for patients with abnormal ALT levels was 0.77 (0.70-0.84). We further performed a sensitivity analysis including 97 NAFLD patients with satisfactory liver biopsy samples that contained at least 11 portal tracts on liver histology (with a specimen length of at least 20 mm). The AUROC for this subset of patients was 0.85 (0.77-0.93), and the Hosmer-Lemeshow test yielded non-significant statistics (p = 0.761), which suggests that there was little departure from perfect fit.

Discussion

To our knowledge, this is the first study that has applied QUS methodology to diagnose NASH in a cohort of well-phenotyped patients with NAFLD. Accordingly, our novel quantitative ultrasound score (the QUS-score), which was derived from the

RF data, had good discriminatory power and calibration capacity for identifying NASH in both the training and validation sets.

iLivTouch is a reliable diagnostic tool for hepatic steatosis and fibrosis assessment.[14,15] The use of *iLivTouch* could lead to highly cost-effective care pathways while minimizing costs, in comparison to either magnetic resonance elastography (MRE) or liver biopsy. Relatively high costs and limited availability, along with contraindications in patients with claustrophobia or implanted metallic devices, further limit the widespread use of magnetic resonance imaging techniques for screening or monitoring of NAFLD. Concerns of repeated exposure to ionizing radiation hinder the role of computed tomography for continuous screening or monitoring of NAFLD. UAP provides a standardized non-invasive measure of hepatic steatosis, and LSM is one of the most widely validated non-invasive tools to detect significant/advanced fibrosis in NAFLD. However, LSM is of limited value for identifying patients with NASH, who have a high risk of progression to cirrhosis and cardiovascular events. Current serum biomarkers include predictive models that have been used for diagnosing NASH. However, biochemical biomarkers and scoring systems derived from them generally have shown a relatively low sensitivity for diagnosis and monitoring of NAFLD.[22] For example, persistently elevated serum ALT concentrations might indicate disease progression; however, patients with advanced stages of NAFLD may still have fairly normal aminotransferase concentrations. Thus, QUS technique would be a good option for a comprehensive assessment of liver

steatosis, fibrosis and NASH within the same examination.

The term radiomics has attracted increasing attention in recent years, and represents the conversion process of medical images into quantitative features, followed by subsequent data analysis for decision support.[32] Naganawa et al. have shown that non-contrast-enhanced computed tomography texture features may predict the presence of NASH.[33] However, in such study, the AUROC for the validation dataset was only 0.60 and the accuracy was 42%. The QUS technique is another novel quantitative imaging technology, involving spectral-based parameterization and envelope statistics, that has shown potential for improving diagnostic capacity in clinical settings.[19] Jeon et al. has investigated the diagnostic performance of QUS parameters for the assessment of hepatic steatosis in patients with NAFLD using magnetic resonance imaging proton density fat fraction (MRI-PDFF) as the reference standard. These authors found that the results derived from QUS data analysis yielded a good correlation with MRI-PDFF and showed good performance for detecting hepatic steatosis.[34] Similar results were reported by Lin et al.[35] Quantitative ultrasound-based imaging techniques have widely been used to detect hepatic steatosis.[22] However, there are limited studies that have examined whether the QUS technology can also be used to non-invasively diagnose NASH. Tang et al. found that a machine learning approach adding QUS parameters to elastography significantly improves the diagnosis of steatohepatitis and the classification of inflammation in an animal model of NASH.[23] However, these authors did not validate their result in humans. To the

best of our knowledge, this is the first study that has applied QUS methodology to diagnose NASH in a cohort of well-phenotype patients with NAFLD.

In our study, 18 candidate ultrasonographic features derived by QUS technology were reduced to two significant predictors by examining the predictor-outcome association by shrinking the regression coefficients with the LASSO method. This statistical method not only surpasses the method of choosing predictors on the basis of the strength of their univariable association with the outcome measure, but also enables the panel of selected features to be combined into a radiomics signature.[36] In our study, the score derived from QUS data analysis using the LASSO method showed good discriminatory capacity and calibration capability for accurately identifying NASH in patients with NAFLD using liver-biopsy as the 'gold standard' for reference. **Figure 4** shows that although the QUS-score increased significantly across the grades of ballooning and lobular inflammation, the discriminatory ability appeared to be better for the latter. As reported above, the two parameters named P7 and P13 that are included in our QUS-score are mainly related to intensity and scattering of the signal, respectively. *Tang* et al. found that a machine learning approach adding QUS parameters (including intensity and scattering parameters) to elastography significantly improved the diagnosis of steatohepatitis and the classification of inflammation in an animal model of NASH.[23] Inflammation through oedema and infiltration of inflammatory cells might increase the internal pressure and liver stiffness. Therefore, although the detailed mechanisms are still unclear, it is possible

to hypothesize that the selected parameters P7 and P13 that reflect the intensity and scattering of liver ultrasound signal are strongly related to the microstructural changes induced by liver inflammation

That said, we believe that the advent of the *iLivTouch* permits a standardized, and rapid identification of NASH for patients at regular intervals. We have developed a toolbox named *LivQ*-box that can be conveniently used to obtain the features of RF data. Our analysis confirms that the correlation between the QUS-score and NASH is good, and we have also established QUS-score cut-offs to rule in and rule out NASH. In fact, the cut-offs of QUS-score for identifying patients with"no-NASH", "indeterminate NASH" or "definite NASH" showed good sensitivity and specificity. In addition, we also tested the diagnostic performance of the QUS-score in different subgroups of individuals, which suggests good applicability of this QUS-score in different populations.

There are some important limitations to our study that should be mentioned. Firstly, our patients with NAFLD are from one country (China), and of a single ethnic group, and therefore may not accurately represent the characteristics of other ethnic groups. Moreover, the design of or single-center study was retrospective and our sample size was relatively small. Larger studies are needed to further validate the utility of the QUS-score in different ethnic groups. Secondly, we have tested the performance of the QUS-score in subgroup analyses, and found that its diagnostic performance was

diminished in patients with severe obesity. Patients with severe obesity always have a long skin-to-liver capsule distance, which would limit the ultrasound diagnostic performance in these patients in terms of accuracy.[37] Thirdly, multiple simultaneously emerging QUS techniques from different vendors may prohibit widespread-buy in and may limit inter-vendor comparisons. Finally, the median length of our liver biopsies was 15 mm (IQR 15-20) and the mean number of portal tracts was 10 (SD 2.5), which were less than the suggested length and number of portal tracts by Rockey et al.[9] It might partly limit the accuracy of histological assessment of NASH. However, the 2014 RCPath 'Tissue Pathways' guideline defines an adequate liver core biopsy as being more than 15 mm and/or containing at least six portal tracts.[29] Considering existing recommendations on the liver specimen size were established for patients with chronic viral hepatitis, it might be different for NAFLD. Unlike chronic hepatitis, NAFLD lesions tend to have a robust and characteristic lobular systematization that mainly affects the centrilobular zone.[38] Although a 20 mm biopsy is considered to be optimal for assessing and quantifying detailed lesions, a 15 mm biopsy usually provides robust information for a global evaluation.[38] For instance, there are some studies that defined the specimen more than 15 mm and/or containing at least six portal tracts as qualified specimens, which were consistent with our research.[29,39] In addition, in a sensitivity analysis, we found that the QUS-score showed good discrimination and calibration in those patients who had a number of portal areas more than 11 and specimen length longer than 20 mm.

In conclusion, the results of our study shows that the newly developed and validated score, which was based on QUS technology, provides a non-invasive and practical way for accurately identifying the presence of NASH both with, and without early fibrosis when lifestyle intervention is important. The QUS score may also be a useful tool to guide risk stratification of patients with NAFLD. Prospective multicenter validation in different ethnic cohorts of patients with NAFLD is now needed to acquire further evidence to support the use of the QUS score in clinical practice.

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 (Baltimore, Md.).* Jul 2016;64:73-84.
2. Huang DQ, El-Serag HB, Loomba R. Global epidemiology of NAFLD-related HCC: trends, predictions, risk factors and prevention. *Nature reviews. Gastroenterology & hepatology.* Dec 21 2020.
3. Younossi ZM, Stepanova M, Ong J, et al. Nonalcoholic Steatohepatitis Is the Most Rapidly Increasing Indication for Liver Transplantation in the United States. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association.* Mar 2021;19:580-589.e585.
4. Zheng KI, Fan JG, Shi JP, et al. From NAFLD to MAFLD: a "redefining" moment for fatty liver disease. *Chinese medical journal.* Oct 5 2020;133:2271-2273.
5. Byrne CD, Targher G. NAFLD: a multisystem disease. *Journal of hepatology.* Apr 2015;62:S47-64.
6. Byrne CD, Targher G. NAFLD as a driver of chronic kidney disease. *Journal of hepatology.* Apr 2020;72:785-801.
7. Targher G, Byrne CD, Tilg H. NAFLD and increased risk of cardiovascular disease: clinical associations, pathophysiological mechanisms and pharmacological implications. *Gut.* Sep 2020;69:1691-1705.
8. Zhou YJ, Wong VW, Zheng MH. Consensus scoring systems for nonalcoholic fatty liver disease: an unmet clinical need. *Hepatobiliary surgery and nutrition.* Jun 2021;10:388-390.
9. Rockey DC, Caldwell SH, Goodman ZD, Nelson RC, Smith AD. Liver biopsy. *Hepatology (Baltimore, Md.).* Mar 2009;49:1017-1044.
10. Zhang YN, Fowler KJ, Ozturk A, et al. Liver fibrosis imaging: A clinical review of ultrasound and magnetic resonance elastography. *Journal of magnetic resonance*

imaging : JMRI. Jan 2020;51:25-42.
11. Chimoriya R, Piya MK, Simmons D, Ahlenstiel G, Ho V. The Use of Two-Dimensional Shear Wave Elastography in People with Obesity for the Assessment of Liver Fibrosis in Non-Alcoholic Fatty Liver Disease. *Journal of clinical medicine.* Dec 29 2020;10.
12. Yoneda M, Honda Y, Nogami A, Imajo K, Nakajima A. Advances in ultrasound elastography for nonalcoholic fatty liver disease. *Journal of medical ultrasonics (2001).* Oct 2020;47:521-533.
13. Xu Y, Liu Y, Cao Z, et al. Comparison of FibroTouch and FibroScan for staging fibrosis in chronic liver disease: Single-center prospective study. *Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver.* Sep 2019;51:1323-1329.
14. Zhu SH, Zheng KI, Hu DS, et al. Optimal thresholds for ultrasound attenuation parameter in the evaluation of hepatic steatosis severity: evidence from a cohort of patients with biopsy-proven fatty liver disease. *European journal of gastroenterology & hepatology.* Mar 1 2021;33:430-435.
15. Duan WJ, Wang XZ, Ma AL, et al. Multicenter prospective study to validate a new transient elastography device for staging liver fibrosis in patients with chronic hepatitis B. *Journal of digestive diseases.* Sep 2020;21:519-525.
16. Zeng F, Wang H, Li X. The defining of the reference range of liver stiffness and fat attenuation parameter for healthy Chinese children. *European journal of gastroenterology & hepatology.* Aug 10 2020.
17. Tadayyon H, Sannachi L, Sadeghi-Naini A, et al. Quantification of Ultrasonic Scattering Properties of In Vivo Tumor Cell Death in Mouse Models of Breast Cancer. *Translational oncology.* Dec 2015;8:463-473.
18. Fontes-Pereira A, Rosa P, Barboza T, et al. Monitoring bone changes due to calcium, magnesium, and phosphorus loss in rat femurs using Quantitative Ultrasound. *Scientific reports.* Aug 10 2018;8:11963.
19. Oelze ML, Mamou J. Review of Quantitative Ultrasound: Envelope Statistics and Backscatter Coefficient Imaging and Contributions to Diagnostic Ultrasound. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control.* Feb 2016;63:336-351.
20. Ghoshal G, Lavarello RJ, Kemmerer JP, Miller RJ, Oelze ML. Ex vivo study of quantitative ultrasound parameters in fatty rabbit livers. *Ultrasound in medicine & biology.* Dec 2012;38:2238-2248.
21. Lin SC, Heba E, Wolfson T, et al. Noninvasive Diagnosis of Nonalcoholic Fatty Liver Disease and Quantification of Liver Fat Using a New Quantitative Ultrasound Technique. *Clinical Gastroenterology And Hepatology.* Jul 2015;13:1337-+.
22. Pirmoazen AM, Khurana A, El Kaffas A, Kamaya A. Quantitative ultrasound approaches for diagnosis and monitoring hepatic steatosis in nonalcoholic fatty liver disease. *Theranostics.* 2020;10:4277-4289.
23. Tang A, Destrempes F, Kazemirad S, Garcia-Duitama J, Nguyen BN, Cloutier G. Quantitative ultrasound and machine learning for assessment of steatohepatitis in a rat model. *European radiology.* May 2019;29:2175-2184.
24. Sun DQ, Zheng KI, Xu G, et al. PNPLA3 rs738409 is associated with renal glomerular and tubular injury in NAFLD patients with persistently normal ALT levels. *Liver international :*

official journal of the International Association for the Study of the Liver. Jan 2020;40:107-119.
25. Castera L, Friedrich-Rust M, Loomba R. Noninvasive Assessment of Liver Disease in Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology.* Apr 2019;156:1264-1281.e1264.
26. Gao F, Huang JF, Zheng KI, et al. Development and validation of a novel non-invasive test for diagnosing fibrotic non-alcoholic steatohepatitis in patients with biopsy-proven non-alcoholic fatty liver disease. *Journal of gastroenterology and hepatology.* Oct 2020;35:1804-1812.
27. Yang YJ, Jung MH, Jeong SH, Hong YP, Kim YI, An SJ. The Association between Nonalcoholic Fatty Liver Disease and Stroke: Results from the Korean Genome and Epidemiology Study (KoGES). *International journal of environmental research and public health.* Dec 21 2020;17.
28. Alberti KG, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation.* Oct 20 2009;120:1640-1645.
29. Seth S, Forrest EH, Morris JM, et al. Audit of medical (non-targeted) liver biopsy specimen quality, pathology reporting and effect on patient management. *Journal of clinical pathology.* May 26 2021.
30. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology (Baltimore, Md.).* Jun 2005;41:1313-1321.
31. Brunt EM, Kleiner DE, Wilson LA, Belt P, Neuschwander-Tetri BA. Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings. *Hepatology (Baltimore, Md.).* Mar 2011;53:810-820.
32. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology.* Feb 2016;278:563-577.
33. Naganawa S, Enooku K, Tateishi R, et al. Imaging prediction of nonalcoholic steatohepatitis using computed tomography texture analysis. *European radiology.* Jul 2018;28:3050-3058.
34. Jeon SK, Lee JM, Joo I, Park SJ. Quantitative Ultrasound Radiofrequency Data Analysis for the Assessment of Hepatic Steatosis in Nonalcoholic Fatty Liver Disease Using Magnetic Resonance Imaging Proton Density Fat Fraction as the Reference Standard. *Korean journal of radiology.* Mar 9 2021.
35. Lin SC, Heba E, Wolfson T, et al. Noninvasive Diagnosis of Nonalcoholic Fatty Liver Disease and Quantification of Liver Fat Using a New Quantitative Ultrasound Technique. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association.* Jul 2015;13:1337-1345.e1336.
36. Ye Z, Zhu Y, Coffman DL. Variable selection for causal mediation analysis using LASSO-based methods. *Statistical methods in medical research.* Mar 23 2021:962280221997505.
37. de Lédinghen V, Wong VW, Vergniol J, et al. Diagnosis of liver fibrosis and cirrhosis

using liver stiffness measurement: comparison between M and XL probe of FibroScan®. *Journal of hepatology.* Apr 2012;56:833-839.

38. Bedossa P. Diagnosis of non-alcoholic fatty liver disease/non-alcoholic steatohepatitis: Why liver biopsy is essential. *Liver international : official journal of the International Association for the Study of the Liver.* Feb 2018;38 Suppl 1:64-66.
39. Zhou YJ, Gao F, Liu WY, et al. Screening for compensated advanced chronic liver disease using refined Baveno VI elastography cutoffs in Asian patients with nonalcoholic fatty liver disease. *Alimentary pharmacology & therapeutics.* Jun 21 2021.

FIGURE LEGENDS

Figure 1. The schematic diagram of quantitative ultrasound parameters calculation. (A) The QUS transducer captures images and radiofrequency (RF) data by *iLivTouch* system. (B) Analyst draws field of interest for signal processing. (C) QUS parameters, including UAP (P1), LSM (P3), intensity (P2-P10), and scattering characteristic (P11-P18), which were calculated by *iLivTouch* system.

Figure 2. Feature selection using the least absolute shrinkage and selection operator (LASSO) binary logistic regression model. (A) LASSO coefficient profiles of the 18 texture features (P1-P18). A coefficient profile plot was produced against the log lambda sequence. The dotted vertical line was drawn at the value selected using 10-fold cross-validation, where two optimal lambda values (by using the minimum criteria and the 1 standard error of the minimum criteria, respectively) both resulted in 2 non-zero coefficients (P7 and P13). (B) Tuning parameter selection in the LASSO model used 10-fold cross-validation via minimum criteria. Binomial deviances from the LASSO regression cross-validation procedure were plotted as a function of log (Lambda). The y-axis indicates binomial deviances. The lower x-axis indicates the log(Lambda l). Numbers along the upper x-axis represent the average number of predictors. Red dots indicate average deviance values for each model with a given Lambda, and vertical bars through the red dots show the upper and lower values of the deviances. Dotted vertical lines were drawn at the optimal values by using the minimum criteria and the 1 standard error of the minimum criteria, respectively (i.e. 1-SE criteria).

Figure 3. Diagnostic performance of the QUS-score for the diagnosis of NASH. (A) AUROC of the training set; (B) AUROC of the validation set; (C) calibration curve of the training set; and (D) calibration curve of the validation set.

Figure 4. Boxplot of the QUS-score versus histopathological severity: (A) NAS score, (B) steatosis grade, (C) lobular inflammation grade, and (D) ballooning grade.

Supplementary Figure 1. The flowchart of the study.