

ORIGINAL ARTICLE

# Discordant pathological diagnosis of non-alcoholic fatty liver disease: A prospective multicenter study

Takuya Kuwashiro,<sup>\*,†</sup> Hirokazu Takahashi,<sup>\*,†</sup> Hideyuki Hyogo,<sup>‡</sup> Yuji Ogawa,<sup>§</sup> Kento Imajo,<sup>§</sup> Masato Yoneda,<sup>§</sup> Takashi Nakahara,<sup>¶</sup> Satoshi Oeda,<sup>\*</sup> Kenichi Tanaka,<sup>†</sup> Yuichiro Amano,<sup>||</sup> Shinji Ogawa,<sup>||</sup> Atsushi Kawaguchi,<sup>\*\*</sup> Shinichi Aishima,<sup>††</sup> Masayoshi Kage,<sup>‡‡</sup> Kazuaki Chayama,<sup>¶¶</sup> Atsushi Nakajima<sup>‡</sup> and Yuichiro Eguchi<sup>\*</sup>

<sup>\*</sup>Liver Center, Saga University Hospital, <sup>†</sup>Division of Metabolism and Endocrinology, Faculty of Medicine, <sup>\*\*</sup>Section of Clinical Cooperation System, Center for Comprehensive Community Medicine, Faculty of Medicine, <sup>††</sup>Department of Pathology and Microbiology, Faculty of Medicine, Saga University, Saga, <sup>‡</sup>Department of Gastroenterology and Hepatology, JA Hiroshima General Hospital, <sup>¶</sup>Department of Gastroenterology and Metabolism, Institute of Biomedical & Health Sciences, Hiroshima University, Hiroshima, <sup>§</sup>Department of Gastroenterology and Hepatology, Yokohama City University Graduate School of Medicine, Yokohama, <sup>‡‡</sup>Kurume University Research Center for Innovative Cancer Therapy, Kurume and <sup>||</sup>Takeda Pharmaceutical Company, Ltd., Fujisawa, Kanagawa, Japan

## Key words

elastography, noninvasive, observer error, reliability.

Accepted for publication 10 November 2019.

## Correspondence

Yuichiro Eguchi, Liver Center, Saga University Hospital, Nabeshima 5-1-1, Saga City, Saga 849-8501, Japan.  
Email: eguchiyu@cc.saga-u.ac.jp

**Declaration of conflict of interest:** The central pathological diagnosis carried out in this study was financially supported by Takeda Pharmaceutical Company.

**Funding support:** Takeda Pharmaceutical Company

## Abstract

**Background:** Liver biopsy has been the standard procedure for diagnosing and evaluating the severity of non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH); however, interobserver discordance remains a critical issue in its pathological diagnosis.

**Methods and Results:** We examined the concordance rates of pathological scoring and diagnosis between pathologists at individual institutions (local diagnosis) and two central pathologists specialized in liver pathology (central diagnosis). A total of 150 patients with NAFLD underwent prospective liver biopsies. NAFLD activity score (NAS) and fibrosis stage were evaluated, and NASH was determined according to Matteoni's classification. NAS, scores for all NAS components, and fibrosis stage were diagnosed at a lower degree by central compared with local diagnosis. NASH was diagnosed in 34% of the patients according to central pathologists compared with 54% according to local pathologists ( $P < 0.001$ ). The concordance rates for NAS, steatosis, inflammation, ballooning, fibrosis, and NASH diagnosis were 26.7, 62.7, 51.3, 48.7, 43.3, and 50.7%, respectively. The correlation coefficient between local and central diagnoses was the lowest for the scoring of ballooning ( $\rho = 0.218$ ).

**Conclusion:** Concordance rates among pathologists for the evaluation of NAFLD are currently poor, and simple and reliable diagnostic and evaluation criteria are urgently needed to improve the clinical management of NAFLD patients.

## Introduction

Non-alcoholic fatty liver disease (NAFLD) is the most common cause of chronic liver disease worldwide and affects about 25% of the world population.<sup>1</sup> NAFLD is classified as non-alcoholic fatty liver (NAFL) or non-alcoholic steatohepatitis (NASH), of which NASH is pathologically characterized by lobular inflammation and the presence of hepatocellular ballooning with or without fibrosis.<sup>2,3</sup> NAFLD increases overall and liver-related mortality, with NASH considered to be a more progressive disease associated with a greater risk of liver cirrhosis and hepatocellular carcinoma than NAFL.<sup>3–5</sup> Therefore, the differential diagnosis of NASH and NAFL has important implications in terms of the prognosis of NAFLD. However, recent studies demonstrated that fibrosis, rather than other histological features, was indicative of all-cause and disease-specific mortality in patients with NAFLD.<sup>6–9</sup> These studies therefore concluded that the

severity of hepatic fibrosis was the most important pathological finding predicting the clinical outcome of NAFLD, rather than a diagnosis of NASH, which requires hepatocyte ballooning according to Matteoni's classification.<sup>2</sup>

Although histological diagnosis by liver biopsy remains the standard procedure for the diagnosis of NASH,<sup>10</sup> liver biopsy is associated with numerous problems, including invasiveness, cost of diagnosis, sampling error, and diagnostic variation among observers.<sup>11–14</sup> However, numerous noninvasive biomarkers and procedures have recently been developed and evaluated for identifying NASH and determining the severity of hepatic fibrosis in patients with NAFLD.<sup>15</sup> In this context, it is necessary to reconsider the significance of liver biopsy for the diagnosis and management of NAFLD and to evaluate the reliability of the pathological diagnosis of NAFLD/NASH and the assessment of its severity among pathologists. We therefore conducted a

prospective multicenter study to compare the diagnostic performances of local and central pathologists for NAFLD.

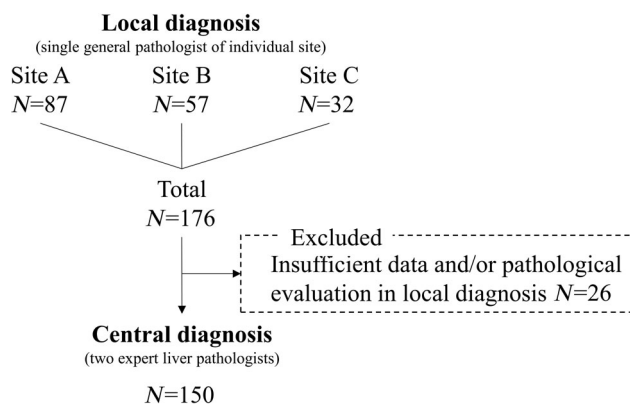
## Methods

**Patients.** A total of 176 consecutive patients with clinically and pathologically diagnosed NAFLD were enrolled from three institutions (Hiroshima University Hospital, Yokohama City University Hospital, and Saga University Hospital) between 2014 and 2016. This cohort was part of the clinical Comprehensive Analysis Study of NAFLD (COMPAS NAFLD; UMIN Clinical Trial Registry UMIN000013323). No patient had any etiology indicative of other liver diseases, including habitual alcohol intake (weekly ethanol consumption >140 g or daily ethanol consumption >20 g), hepatitis B surface antigen or hepatitis C antibody positivity, or abnormal serum thyroid hormone levels, and no patient had autoimmune liver disease, drug-induced hepatotoxicity, hemochromatosis, or Wilson's disease. The protocol was approved by the clinical research ethics review committee of each facility. Each patient gave written informed consent to participate in the clinical study. The ethics committee of each participating institution approved this study, which was performed in accordance with the principles of the 1975 Declaration of Helsinki, revised in 2013.

### Liver biopsy procedure and pathological evaluation.

All patients underwent liver biopsy, and liver specimens were obtained percutaneously using a 16-gauge biopsy needle. Biopsies were performed in the right lobe of the liver under ultrasound guidance. All liver biopsy samples were at least 20 mm long. The specimens were fixed in 10% formalin, paraffin embedded, sectioned, and stained with hematoxylin–eosin and Azan or Masson-trichrome stain for histological evaluation. All liver biopsy specimens were evaluated by a single experienced general pathologist in the local institution, who was unaware of the clinical conditions and patient data and gave a local diagnosis. The same slides were also evaluated simultaneously by two different pathologists designated by our study group (S.A. and M.K.), who were experts in liver pathology and gave a consensual central diagnosis. All pathologists were unaware of the clinical information and were certified by the Japanese Society of Pathology. The samples were scored according to the NAFLD activity score (NAS).<sup>16</sup> Lobular inflammation and steatosis were scored on a scale of 0–3 and ballooning on a scale of 0–2. Fibrosis stage was scored on a 5-point scale (F0–F4) according to the Kleiner classification.<sup>16</sup> NAFLD was defined as excessive fat accumulation in the liver with more than 5% of hepatocytes.<sup>10</sup> NASH diagnosis was based on Matteoni's classification.<sup>2</sup> In the 176 samples evaluated by the central pathologist, 26 samples were excluded because of insufficient data and/or pathological evaluation in local diagnosis. The scoring, classification, and NASH diagnosis were compared between the local diagnosis and central diagnosis (Fig. 1). Concordance rate was calculated as the ratio of the total number of the samples ( $n = 150$ ) to the number of the samples with concordant score or diagnosis between the central diagnosis and local diagnosis.

**Statistical analysis.** The local and central diagnostic assessments were compared using Wilcoxon's signed-rank test, McNemar's test, and Spearman's rank correlation coefficient



**Figure 1** Study design. A single general pathologist evaluated liver biopsy in the individual site. All samples were collected and evaluated by the two expert central liver pathologists.

analysis. Agreement of ordered categorical variables was evaluated by weighted  $\kappa$  with linear weighting, which considers the proximity of categories.<sup>17</sup> All statistical analyses were carried out using IBM SPSS Statistics ver.25 software (SPSS Japan, Tokyo, Japan).

## Results

### Comparison of pathological scoring and staging of NAFLD between local and central pathologists.

The pathological diagnoses made by the local and central pathologists are shown in Table 1 and Figure 2. NAS, individual components of NAS, and fibrosis stage were diagnosed at a significantly lower degree by the central, compared with the local, pathologists. Eighty-one patients (54%) received a diagnosis of NASH based on Matteoni's classification, according to local diagnosis, compared with only 51 patients (34%) according to central diagnosis. Central pathologists diagnosed eight patients with non-NAFLD according to Matteoni's classification.

### Concordance between local and central diagnoses.

The concordance rates for the diagnosis of steatosis, lobular inflammation, and ballooning according to NAS were 62.7, 51.3,

**Table 1** Comparison of pathological scoring and staging of NAFLD between local and central pathologists

	Local ( $n = 150$ )	Central ( $n = 150$ )	<i>P</i> value
NAS (0–4/5–8)	105/45	132/18	<0.001
Steatosis (0/1/2/3)	0/67/59/24	8/80/36/26	<0.001
Lobular inflammation (0/1/2/3)	3/100/40/7	31/105/13/1	<0.001
Ballooning (0/1/2)	69/54/27	94/48/8	<0.001
Fibrosis stage (F0/1/2/3/4)	13/66/33/32/6	45/55/30/16/4	<0.001
Matteoni's classification (non-NAFLD/Type1-2/ Type3-4)	0/69/81	8/91/51	<0.001 <sup>†</sup>

<sup>†</sup>Comparison in distribution of NASH or non-NASH by McNemar's test.

and 48.7%, respectively (Table 2), with concordances of 26.7% for total NAS and 43.3% for fibrosis stage. The diagnosis concordance for distinguishing between NAFL and NASH was 50.7% according to Matteoni's classification. The  $\kappa$  score was also determined to evaluate diagnosis concordance (Table 3). Similar to the result of concordance rates,  $\kappa$  score for the diagnosis of ballooning (0.57) was lower than that for other components of NAS and fibrosis stage.

**Correlation between local and central diagnoses.**

We tested the correlations between the local and central diagnoses and compared the correlation coefficients among the different pathological findings. There was a significant correlation between the local and central diagnoses for steatosis ( $\rho = 0.709$ ,  $P < 0.0001$ ) but lower correlation coefficients for inflammation ( $\rho = 0.286$ ,  $P = 0.0005$ ) and ballooning ( $\rho = 0.218$ ,  $P = 0.0079$ ) (Fig. 3a–c). Correlation for fibrosis stage was the most significant pathological finding ( $\rho = 0.627$ ,  $P < 0.0001$ ) (Fig. 3d).

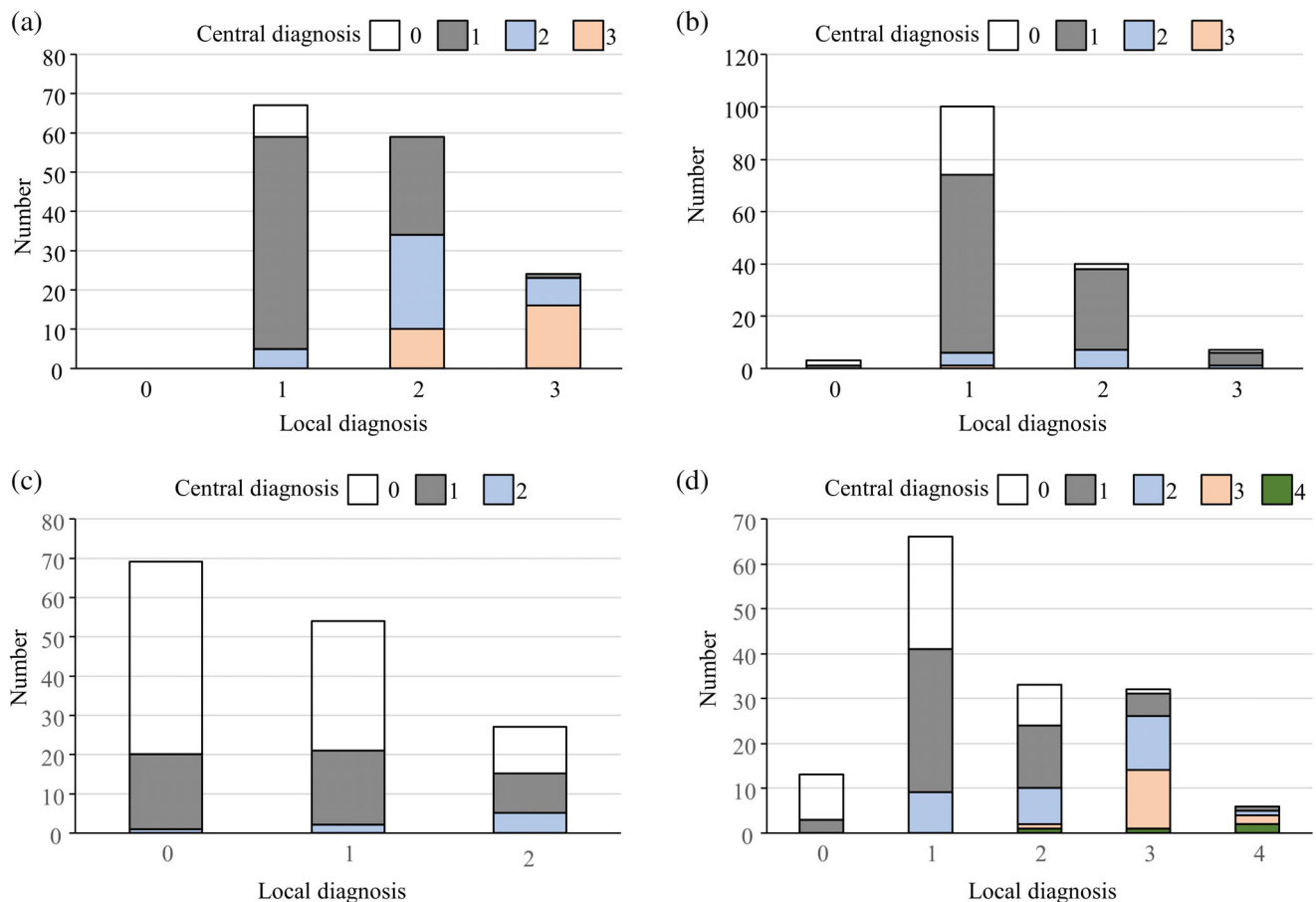
**Table 2** Diagnosis concordance rate between local and central pathologists

	Concordance rate (%) <sup>†</sup>
NAS	26.7
Steatosis	62.7
Lobular inflammation	51.3
Ballooning	48.7
Fibrosis	43.3
NASH diagnosis	50.7

<sup>†</sup>Concordance rate was calculated as the ratio of the total number of the samples ( $n = 150$ ) to the number of the samples with concordant score or diagnosis between central diagnosis and local diagnosis.

**Discussion**

The current study demonstrated a serious discordance between pathologists in the pathological diagnosis of NAFLD. Focusing on the individual pathological components, the lowest concordance (48.7%) and lowest correlation coefficient ( $\rho = 0.218$ ) between



**Figure 2** Discordance of pathological evaluation of NAFLD between local and central pathologists. (a) steatosis, (b) lobular inflammation, (c) ballooning, and (d) fibrosis

**Table 3** Diagnosis agreement between local and central pathologists

Categories (diagnosis)	$\kappa$ score
Steatosis (score 0–1 or score 2–3 in NAS)	0.79
Lobular inflammation (score 0–1 or score 2–3 in NAS)	0.70
Ballooning (score 0 or score 1–2 in NAS)	0.57
Fibrosis (stages 0–1 or stages 2–4 in Kleiner classification)	0.74
NASH diagnosis (non-NASH or NASH in Matteoni's classification)	0.53

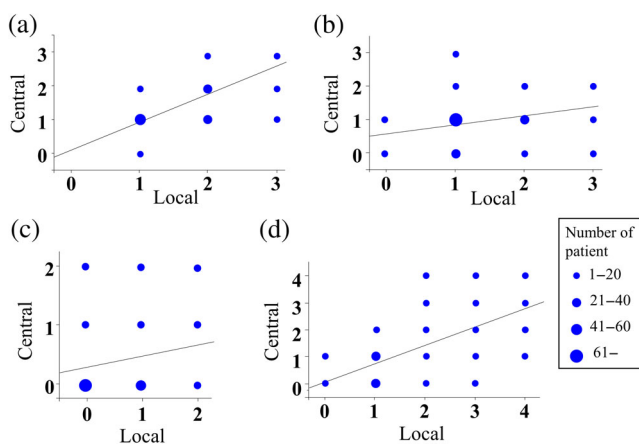
local and central diagnoses was observed for the diagnosis of ballooning. These data suggest the existence of significant interobserver error in the diagnosis of hepatocyte ballooning, which is a key finding for the diagnosis of NASH.<sup>2,10,16</sup> Local pathologists identified ballooning (grades 1–2) in 54% of patients, compared with only 37.3% according to central diagnosis, suggesting that pathologists not specialized in liver pathology might overdiagnose ballooning. The diagnosis of inflammation showed a similar trend, with 2% of patients diagnosed with grade 0 inflammation according to local diagnosis compared with almost 10 times more patients according to the central diagnosis (20.7%). These discordances indicated poor reliability for a diagnosis of NASH according to Matteoni's classification.

Hepatic fibrosis has been strongly implicated in the long-term prognosis of NAFLD patients.<sup>6–9</sup> Moreover, NAFLD prognosis is independent of the diagnosis of NASH/non-NASH.<sup>7,8</sup> It is therefore critical to identify patients at higher risk of NAFLD with advanced fibrosis in order to optimize their management. However, our results showed that the concordance rate for a diagnosis of fibrosis was only 43.3%. Moreover, local pathologists diagnosed stage 3 or 4 fibrosis in 25.3% of patients, compared with 13.3% by central pathologists. Previous studies showed that the liver-related mortality rate increased exponentially in NAFLD patients with advanced liver fibrosis.<sup>8</sup> An

overdiagnosis of hepatic fibrosis would increase the number of patients with therapeutic indications, thus increasing the economic burden in light of the upcoming availability of novel therapeutic agents for fibrosis in NAFLD.

Numerous studies have investigated the interobserver reliability for pathological diagnosis. Theodossi *et al.* reported a concordance rate of final pathological diagnosis of 15% in various liver diseases.<sup>17</sup> In chronic viral hepatitis, agreement of scoring and staging between the general pathologists and expert pathologists was not adequate, and the experience level of pathologists could affect the agreement of diagnosis.<sup>18,19</sup> In NAFLD, Younossi *et al.* identified that discrepancy in the diagnosis of inflammation among pathologists is more severe in comparison with fibrosis.<sup>13</sup> Juluri *et al.* reported that agreement ( $\kappa$  score) between the diagnosis of two pathologists (community pathologist and expert pathologist) was 0.62 for steatosis, 0.44 for lobular inflammation, 0.25 for ballooning, 0.40 for NAS, 0.35 for fibrosis, and 0.46 for non-NASH/NASH diagnosis,<sup>20</sup> suggesting that interobserver reliability was the highest in the diagnosis of steatosis and lowest in the diagnosis of ballooning. Our study confirmed the findings of Juluri *et al.* that agreement on the diagnosis of ballooning is the most difficult to obtain. In contrast, however, the  $\kappa$  score for the diagnosis of fibrosis stage in the study by Juluri *et al.* was only 0.35 and was lower than for other pathological diagnoses, whereas our study showed better agreement ( $\kappa = 0.76$  in Table 2). Gawrieh *et al.* compared the diagnosis of one senior pathologist with that of one junior pathologist and demonstrated  $\kappa = 0.72$  for the diagnosis of steatosis,  $\kappa = 0.64$  for the diagnosis of fibrosis stage, and  $\kappa = 0.32$  for the diagnosis of ballooning.<sup>14</sup> According to this evidence and the results of our study, concordance in the diagnosis of ballooning tends to be low for NAFLD, whereas concordance in the diagnosis of fibrosis stage varies among studies. Moreover, differences in experience between the pathologists could be a factor that affects concordance. Indeed, training and prior consent of scoring and diagnosis, including definitions of detailed morphological criteria, increase the concordance of diagnosis and scoring by pathologists.<sup>14,21</sup>

The  $\kappa$  score obtained in the current study was 0.53–0.79 for the comparison between central diagnosis and local diagnoses, which could be interpreted as “moderate” or “good.”<sup>22</sup> However, there are several limitations in an evaluation of interobserver reliability using the  $\kappa$  score. It was reported that prevalence bias could affect  $\kappa$  score.<sup>23</sup> For example, a significant difference in the prevalence among the categories could result in either a significantly high or low  $\kappa$  score. Moreover, because the scoring is quantitative, a weighted  $\kappa$  score, which generally tends to be higher than a non-weighted score, should be statistically used as we did in the current study; however, in terms of clinical significance, a difference of 1 point in the scoring system creates a serious discrepancy. For example, the difference between a ballooning score of 0 and 1 could result in a diagnosis of non-NASH or NASH. Therefore, the nonweighted  $\kappa$  score could also be referenced for the evaluation of concordance of pathological diagnosis to determine the difference between clinically important scores. Indeed, concordance was poor to moderate if a nonweighted  $\kappa$  score was used in the current study ( $\kappa = 0.59$  for steatosis score 0–1 or 2–3,  $\kappa = 0.14$  for lobular inflammation score 0–1 or 2–3,  $\kappa = 0.15$  for ballooning score 0 or 1–2, and  $\kappa = 0.47$  for fibrosis stage 0–1 or 2–4;



**Figure 3** Correlation analysis between local and central diagnosis in (a) steatosis, (b) lobular inflammation, (c) hepatocyte ballooning, and (d) fibrosis. Dot size represents the number of patients in each score.

data not shown). Taken together, because the  $\kappa$  score could mislead the interpretation of agreement, the results of simple concordance/discordance rate, correlation coefficient, and clinical significance should be carefully considered for fuller comprehension of agreement.<sup>23</sup>

Liver biopsy remains the gold standard for characterizing changes in liver histology in patients with NAFLD; however, liver biopsy has some limitations, including its cost, the risks of morbidity and (rarely) mortality, and the need for adequate experience to provide a pathological diagnosis.<sup>10</sup> It has therefore been considered that liver biopsy should only be performed in patients most likely to benefit from the diagnosis, therapeutic guidance, and prognostic information.<sup>24</sup> However, the pathological diagnosis of NAFLD should now be reconsidered in terms of NASH, which does not affect the clinical outcome, and in terms of the evaluation of histological fibrosis, which is the most important finding for predicting mortality risk in NAFLD. Novel reliable screening and diagnostic strategies based on the evaluation of NAFLD and fibrosis, other than liver biopsy, are thus required to identify NAFLD patients at significant risk of mortality.<sup>25</sup>

Recent research into noninvasive biomarkers for detecting hepatic fibrosis in NAFLD may affect the clinical significance of liver biopsy. Serum biomarkers, including keratin 18,<sup>26,27</sup> type III procollagen peptide,<sup>28</sup> and type III collagen propeptide,<sup>29</sup> as well as noninvasive scoring systems, including the NAFLD fibrosis score,<sup>30</sup> Fibrosis-4 index,<sup>31</sup> AST/platelet ratio index,<sup>32</sup> FibroMeter,<sup>33,34</sup> and BARD score,<sup>35</sup> have demonstrated high diagnostic accuracy and reliability for evaluating liver fibrosis and diagnosing NASH. Ultrasound, including vibration-controlled transient elastography (VCTE), has also been well studied and has been used clinically to predict liver fibrosis in NAFLD. Two recent studies demonstrated the excellent diagnostic performance of VCTE in patients with biopsy-proven NAFLD.<sup>36,37</sup> Furthermore, the imaging-based magnetic resonance (MR) technique, MR elastography, showed greater diagnostic accuracy than VCTE for the prediction of liver fibrosis in patients with NAFLD,<sup>38,39</sup> as recently confirmed in a multicenter study.<sup>40</sup> As an alternative to liver biopsy, MR elastography could serve as a new gold standard for the assessment of liver fibrosis with NAFLD.<sup>41</sup> The rate of agreement on fibrosis stage between different radiologists reading MR elastography has been shown to be greater than that of separate pathologists assessing biopsy specimens.<sup>32,33,42–44</sup> Therefore, in our opinion, MR elastography will become a new gold standard and benchmark, which should be used to evaluate the utility of biomarkers. MR imaging can also be used for the noninvasive quantification of liver steatosis by spectroscopy<sup>45</sup> or by measuring the proton density fat fraction,<sup>46,47</sup> as widely used in NAFLD clinical trials.<sup>48</sup> Overall, therefore, several noninvasive procedures are available for predicting liver fibrosis in NAFLD. Although further studies are needed to determine whether these procedures predict clinical outcomes, including mortality, in patients with NAFLD, they may provide alternatives to liver biopsy in clinical practice.

In conclusion, there is significant discordance among pathologists in relation to the diagnosis of NASH and in NAS scoring. Discordances in the diagnosis of ballooning and fibrosis are critical for the diagnosis of NASH and the management of NAFLD, respectively. There is therefore an urgent need for globally agreed, simple pathological diagnostic criteria for NAFLD

or the establishment of alternative noninvasive and quantitative methods to provide diagnostic and reference information for the clinical management of NAFLD patients.

## Acknowledgments

We thank all the medical staff of Saga University Hospital, Yokohama City University Hospital, and Yokohama JA Hiroshima General Hospital. We also thank Susan Furness, PhD and Hugh McGonigle from the Edanz Group ([www.edanzediting.com/ac](http://www.edanzediting.com/ac)) for editing drafts of the manuscript.

## 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. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. *Gastroenterology*. 1999; **116**: 1413–19.
- 3 Chalasani N, Younossi Z, Lavine JE *et al.* The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. *Hepatology*. 2012; **55**: 2005–23.
- 4 Ekstedt M, Franzén LE, Mathiesen UL *et al.* Long-term follow-up of patients with NAFLD and elevated liver enzymes. *Hepatology*. 2006; **44**: 865–73.
- 5 Söderberg C, Stål P, Askling J *et al.* Decreased survival of subjects with elevated liver function tests during a 28-year follow-up. *Hepatology*. 2010; **51**: 595–602.
- 6 Ekstedt M, Hagström H, Nasr P *et al.* Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. *Hepatology*. 2015; **61**: 1547–54.
- 7 Angulo P, Kleiner DE, Dam-Larsen S *et al.* Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2015; **149**: 389–97.
- 8 Dulai PS, Singh S, Patel J *et al.* Increased risk of mortality by fibrosis stage in non-alcoholic fatty liver disease: systematic review and meta-analysis. *Hepatology*. 2017; **65**: 1557–65.
- 9 Hagström H, Nasr P, Ekstedt M *et al.* Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. *J. Hepatol.* 2017; **67**: 1265–73.
- 10 Chalasani N, Younossi Z, Lavine JE *et al.* The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. *Hepatology*. 2018; **67**: 328–57.
- 11 Ratziu V, Charlotte F, Heurtier A *et al.* Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology*. 2005; **128**: 1898–906.
- 12 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.
- 13 Younossi ZM, Gramlich T, Liu YC *et al.* Nonalcoholic fatty liver disease: assessment of variability in pathologic interpretations. *Mod. Pathol.* 1998; **11**: 560–5.
- 14 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.

- 15 Vilar-Gomez E, Chalasani N. Non-invasive assessment of non-alcoholic fatty liver disease: Clinical prediction rules and blood-based biomarkers. *J. Hepatol.* 2018; **68**: 305–15.
- 16 Kleiner DE, Brunt EM, Van Natta M *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology.* 2005; **41**: 1313–21.
- 17 Theodossi A, Skene AM, Portmann B *et al.* Observer variation in assessment of liver biopsies including analysis by kappa statistics. *Gastroenterology.* 1980; **79**: 232–41.
- 18 Rousselet MC, Michalak S, Dupre F *et al.* Sources of variability in histological scoring of chronic viral hepatitis. *Hepatology.* 2005; **41**: 257–64.
- 19 Woynarowski M, Cielecka-Kuszyk J, Kaluzynski A *et al.* Inter-observer variability in histopathological assessment of liver biopsies taken in a pediatric open label therapeutic program for chronic HBV infection treatment. *World J. Gastroenterol.* 2006; **12**: 1713–17.
- 20 Juluri R, Vuppalanchi R, Olson J *et al.* Generalizability of the Non-alcoholic Steatohepatitis Clinical Research Network histologic scoring system for nonalcoholic fatty liver disease. *J. Clin. Gastroenterol.* 2011; **45**: 55–8.
- 21 Jung ES, Lee K, Yu E *et al.* Interobserver agreement on pathologic features of liver biopsy tissue in patients with nonalcoholic fatty liver disease. *J. Pathol. Transl. Med.* 2016; **50**: 190–6.
- 22 Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*, 3rd edn. New Jersey, USA: John Wiley & Sons, Inc., 2003.
- 23 Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J. Clin. Epidemiol.* 1993; **46**: 423–9.
- 24 Rinella ME, Sanyal AJ. Management of NAFLD: a stage-based approach. *Nat. Rev. Gastroenterol. Hepatol.* 2016; **13**: 196–205.
- 25 Yoneda M, Imajo K, Takahashi H *et al.* Clinical strategy of diagnosing and following patients with nonalcoholic fatty liver disease based on invasive and noninvasive methods. *J. Gastroenterol.* 2018; **53**: 181–96.
- 26 Kwok R, Tse YK, Wong GL *et al.* Systematic review with meta-analysis: non-invasive assessment of non-alcoholic fatty liver disease—the role of transient elastography and plasma cytokeratin-18 fragments. *Aliment. Pharmacol. Ther.* 2014; **39**: 254–69.
- 27 Feldstein AE, Wieckowska A, Lopez AR, Liu YC, Zein NN, McCullough AJ. Cytokeratin-18 fragment levels as noninvasive biomarkers for nonalcoholic steatohepatitis: a multicenter validation study. *Hepatology.* 2009; **50**: 1072–8.
- 28 Tanwar S, Trembling PM, Guha IN *et al.* Validation of terminal peptide of procollagen III for the detection and assessment of non-alcoholic steatohepatitis in patients with nonalcoholic fatty liver disease. *Hepatology.* 2013; **57**: 103–11.
- 29 Nielsen MJ, Veidal SS, Karsdal MA *et al.* Plasma Pro-C3 (N-terminal type III collagen propeptide) predicts fibrosis progression in patients with chronic hepatitis C. *Liver Int.* 2015; **35**: 429–37.
- 30 Angulo P, Hui JM, Marchesini G *et al.* The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology.* 2007; **45**: 846–54.
- 31 Sterling RK, Lissen E, Clumeck N *et al.* Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology.* 2006; **43**: 1317–25.
- 32 Lin ZH, Xin YN, Dong QJ *et al.* Performance of the aspartate aminotransferase-to-platelet ratio index for the staging of hepatitis C-related fibrosis: an updated meta-analysis. *Hepatology.* 2011; **53**: 726–36.
- 33 Leroy V, Sturm N, Faure P *et al.* Prospective evaluation of FibroTest(R), FibroMeter(R), and HepaScore(R) for staging liver fibrosis in chronic hepatitis B: comparison with hepatitis C. *J. Hepatol.* 2014; **61**: 28–34.
- 34 Rosenberg WM, Voelker M, Thiel R *et al.* Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology.* 2004; **127**: 1704–13.
- 35 Harrison SA, Oliver D, Arnold HL, Gogia S, Neuschwander-Tetri BA. Development and validation of a simple NAFLD clinical scoring system for identifying patients without advanced disease. *Gut.* 2008; **57**: 1441–7.
- 36 Tapper EB, Challies T, Nasser I, Afdhal NH, Lai M. The performance of vibration controlled transient elastography in a US cohort of patients with nonalcoholic fatty liver disease. *Am. J. Gastroenterol.* 2016; **111**: 677–84.
- 37 Nozaki Y, Fujita K, Wada K *et al.* Deficiency of eNOS exacerbates early-stage NAFLD pathogenesis by changing the fat distribution. *BMC Gastroenterol.* 2015; **15**: 177.
- 38 Imajo K, Kessoku T, Honda Y *et al.* Magnetic resonance imaging more accurately classifies steatosis and fibrosis in patients with non-alcoholic fatty liver disease than transient elastography. *Gastroenterology.* 2016; **150**: 626–37.
- 39 Looma R, Wolfson T, Ang B *et al.* Magnetic resonance elastography predicts advanced fibrosis in patients with nonalcoholic fatty liver disease: a prospective study. *Hepatology.* 2014; **60**: 1920–8.
- 40 Hsu C, Caussy C, Imajo 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.
- 41 Mumtaz S, Schomaker N, Von Roenn N. Pro: noninvasive imaging has replaced biopsy as the gold standard in the evaluation of non-alcoholic fatty liver disease. *Clin. Liver Dis. (Hoboken).* 2019; **13**: 111–13.
- 42 Yoon JH, Lee JM, Joo I *et al.* Hepatic fibrosis: prospective comparison of MR elastography and US shear-wave elastography for evaluation. *Radiology.* 2014; **273**: 772–82.
- 43 Rustogi R, Horowitz J, Harmath C *et al.* Accuracy of MR elastography and anatomic MR imaging features in the diagnosis of severe hepatic fibrosis and cirrhosis. *J. Magn. Reson. Imaging.* 2012; **35**: 1356–64.
- 44 Pavlides M, Birks J, Fryer E *et al.* Interobserver variability in histologic evaluation of liver fibrosis using categorical and quantitative scores. *Am. J. Clin. Pathol.* 2017; **147**: 364–9.
- 45 Reeder SB, Cruite I, Hamilton G, Sirlin CB. Quantitative assessment of liver fat with magnetic resonance imaging and spectroscopy. *J. Magn. Reson. Imaging.* 2011; **34**: 729–49.
- 46 Idilman IS, Keskin O, Celik A *et al.* A comparison of liver fat content as determined by magnetic resonance imaging-proton density fat fraction and MRS versus liver histology in non-alcoholic fatty liver disease. *Acta Radiol.* 2016; **57**: 271–8.
- 47 Heba ER, Desai A, Zand KA *et al.* Accuracy and the effect of possible subject-based confounders of magnitude-based MRI for estimating hepatic proton density fat fraction in adults, using MR spectroscopy as reference. *J. Magn. Reson. Imaging.* 2016; **43**: 398–406.
- 48 Nouredin M, Lam J, Peterson MR *et al.* Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials. *Hepatology.* 2013; **58**: 1930–40.