


# Poor Inter-test Reliability Between CK18 Kits as a Biomarker of NASH

Carolina F. M. G. Pimentel<sup>1</sup>  · Zhenghui G. Jiang<sup>1</sup> · Takeshi Otsubo<sup>1</sup> · Linda Feldbrugge<sup>1</sup> · Tracy L. Challies<sup>2</sup> · Imad Nasser<sup>2</sup> · Simon Robson<sup>1</sup> · Nezam Afdhal<sup>1</sup> · Michelle Lai<sup>1</sup>

Received: 4 August 2015 / Accepted: 2 October 2015 / Published online: 13 October 2015  
© Springer Science+Business Media New York 2015

## Abstract

**Background and Aim** Nonalcoholic fatty liver disease (NAFLD) affects 15–40 % of the general population; 10–20 % of those patients have a more severe form of the disease known as nonalcoholic steatohepatitis (NASH). Cytokeratin-18 (CK18), released during apoptosis and one of the most studied biomarkers in NASH, can be measured by a number of commercially available kits. We compared serum measurements of the CK18 M30 from two different kits using the same cohort to evaluate the reliability between two test kits. **Methods** We measured serum levels of CK18 M30 from 185 patients with biopsy-proven NAFLD from a single center from 2009 to 2015, using two different ELISA kits, Test 1 (T1) and Test 2 (T2). Advanced fibrosis was defined as fibrosis stages 3–4 and NASH defined by NAS score  $\geq 5$ . **Results** Mean age was 50.2 years (SD 12.6), 61.1 % male and 87 % White; 49.6 % had NASH and 32.2 % advanced fibrosis. There was no significant correlation between measurements from the two kits ( $p = 0.86$ ,  $r = 0.01$ ). While T2 predicted NASH and advanced fibrosis, T1 did not. The area under ROC curve for the prediction of NASH was 0.631 for T2 versus 0.500 for T1.

**Conclusions** Measurements from two different CK18 M30 test kits did not correlate with each other. One kit showed statistically significantly higher levels of CK18

M30 in patients with advanced fibrosis and NASH, while the other kit did not. With the increasing use of CK18 as a biomarker in NASH, it is important to standardize the different kits as it could greatly bias the results.

**Keywords** Cytokeratin 18 · Nonalcoholic liver disease · Test reliability · Liver fibrosis · Apoptosis · Biomarker

## Abbreviations

NAFLD	Nonalcoholic liver disease
NASH	Nonalcoholic steatohepatitis
CK18	Cytokeratin 18
AUROC	Area under the receiver operating characteristic curve
NAFL	Nonalcoholic fatty liver, simple steatosis
HR	Hazard ratio

## Introduction

Nonalcoholic fatty liver disease (NAFLD) is the most common cause of liver disease worldwide, affecting approximately 20 % of the general population, based on different assessment methods such as ultrasonography, liver biopsy, or liver test abnormalities [1]. In the USA, some studies based on MRI evaluation estimate a prevalence of 31 %, while others 47 % based on ultrasonography findings [2, 3]. While the prevalence of other liver diseases has been stable over the last 20 years, NAFLD has increased remarkably over the last years along with the prevalence of metabolic syndrome [4]. With the current and rising prevalence of NAFLD, diagnostic tools not only for disease assessment but also for follow-up are required.

✉ Carolina F. M. G. Pimentel  
carolinapimentel.gastro@gmail.com

<sup>1</sup> Division of Gastroenterology/Liver Center, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, 110 Francis Street, Suite 4 A, Boston, MA 02215, USA

<sup>2</sup> Division of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

The natural history of NAFLD depends on severity of the liver damage and population risk factors, such as obesity and diabetes [5]. Disease spectrum ranges from the mild benign form with simple steatosis (NAFL) to the more severe form of the disease in 10–20 % of these patients, known as nonalcoholic steatohepatitis (NASH), characterized by steatosis, inflammation, and hepatocyte ballooning (apoptosis), leading to fibrosis [1, 6]. NASH has a more severe disease course, in which patients are more susceptible to develop liver cirrhosis and hepatocellular carcinoma [7, 8], so identifying patients at risk of end-stage liver disease is essential during evaluation and management [9]. In addition, patients with NASH have increased overall mortality, cardiovascular diseases, and liver-related complications [10]. Although some studies suggest that not only NASH but also NAFL can progress to liver fibrosis, the annual progression rate in patients with steatohepatitis in liver biopsy is higher than simple steatosis (one stage of progression over 7.1 vs. 14.3 years, respectively) [11–13]. Moreover, the prognosis of patients with NASH appears to be directly related to fibrosis stages, which can predict both overall and disease-specific mortality. Ekstedt et al. [14] demonstrated that patients with fibrosis 3–4, irrespective of NAS, had increased mortality, with a HR 3.3.

At present, liver biopsy is the current way to evaluate the presence of NASH and assess the severity of liver damage; however, it is an invasive and expensive procedure, with known sampling variability and risks of complications such as bleeding, bile leaks, and death is rare cases. Therefore, noninvasive methods to distinguish between NAFLD and NASH and to predict fibrosis stage are imperatively required.

A number of serum biomarkers have been studied in NASH, with the most well-studied biomarker being cytokeratin-18 (CK18) fragments. CK18 is a keratin-containing protein involved in cytoskeleton cell formation. During hepatocyte apoptosis, it is cleaved by caspases and released in circulation as fragments (M30 and M65), which are easily detectable and used as markers of apoptosis to identify patients with NASH. CK18 M30 fragment is more related to apoptosis than M65, as the latter can be detected during tissue necrosis, and may be released in circulation despite caspase activation [15].

The first study addressing its importance was presented by Wieckowska et al. [16] when a pilot study demonstrated that caspase-generated CK18 fragments were higher in patients with NAFLD compared to controls and able to differentiate between low and high fibrosis and inflammation stages. Since then, numerous studies suggest a wide range of potential usefulness of this test as a noninvasive tool in NASH diagnosis, to differentiate NASH from simple steatosis and as a marker of disease severity and response to treatment [17–19]. In addition, recent studies

suggest models combining CK18 with different biomarkers to improve diagnostic performance [20, 21].

Although previous studies [1, 22] showed that CK18 is a clinical relevant biomarker and has sufficient diagnostic accuracy in detection or exclusion of NASH, recent publications showed poorer test performance of CK18 with a low sensitivity as a diagnostic test for NASH [23, 24]. The differences in test performance are seen between studies using the same ELISA kits and between studies that use different kits. There are a number of CK18 fragment ELISA kits commercially available. However, there does not seem to be standardization between the kits, or data available comparing the test performance of different kits. Different studies on the test performance of CK18 as a NASH biomarker have varying results, raising the question of whether the different findings are due in part to differences in the test kits.

The aim of this study is to compare serum measurements of the CK18 M30 fragments by two different CK18 M30 ELISA kits using the same cohort of patients to test inter-test reliability.

## Materials and Methods

### Patient Population

The subjects for this study were enrolled in an NAFLD patient registry at Beth Israel Deaconess Medical Center (BIDMC) from 2009 to 2014. The BIDMC NAFLD patient registry is a prospective study of subjects with biopsy-proven NAFLD. Patients with other chronic liver diseases or consumption of >20 g alcohol daily were excluded from the registry. Data on patient demographics, medical history, and from their physical examination were obtained at the enrollment of the study. At the time of this study, there were 183 patients enrolled in the NAFLD registry study. The study has been approved by the BIDMC institutional review board.

### Liver Biopsy and Histological Assessment

Ultrasound-guided liver biopsy was performed within 3 months of the baseline visit and evaluated by an experienced liver pathologist, unaware of subjects' clinical information, using the NASH Clinical Research Network Scoring System developed by Kleiner et al. [25]. Advanced fibrosis and NASH were defined by liver biopsy findings. Advanced fibrosis was defined as fibrosis stages 3–4. According to Kleiner et al. [25] NAFLD activity score (NAS score) was also considered for diagnosis of NASH when >5. We defined severe hepatocyte ballooning as having a hepatocyte ballooning score of 2 and severe lobular inflammation as having score of 2–3. For subsequent analyses, we studied the role of CK18 in prediction

of NASH and advanced fibrosis, as these are the most important clinical predictors related to disease severity and mortality [11, 13].

### Clinical Biochemistry and Measurements of CK18 Fragments

Laboratory tests and collection of serum were performed at enrollment. Baseline serum was stored at  $-80^{\circ}\text{C}$ . Routine blood tests including complete blood count, chemistry, liver function tests, albumin, and lipid panel were processed at the BIDMC clinical laboratory.

Serum CK18 M30 fragment was measured using two different ELISA kits: Human cytokeratin 18<sup>®</sup>, Biotang (Test 1) and M30-Apoptosense, PEVIVA<sup>®</sup>, DiaPharma (Test 2). All procedures were done according to kit manufacturer's instructions. Intra-assay variability was evaluated by using the intra-assay variability coefficient (CV%) for both tests. CV% was calculated by using the standard deviation of the means after bootstrap analysis from each test, divided by the overall mean of each group. CV% <10 was considered acceptable.

#### Test 1

CK18 concentration in the sera was measured using the Human CK18 ELISA kit (Biotang, Lexington, MA, USA). Antihuman KRT18 mAb is derived from hybridization of mouse FO myeloma cells with spleen cells from BALB/c mice immunized with recombinant human KRT18 amino acids 79-430 purified from *Escherichia coli*. This kit is a presentation of using antibody against CK18 full protein, able to detect all fragments generated after caspase activation. Thus, high levels of this biomarker are directly related to M30 fragment levels, the main fragment released during the apoptosis process.

The ELISA was performed following protocol provided by the manufacturer. In short, 10  $\mu\text{L}$  of serum samples was pipetted into the pre-coated well along with standards. After incubation, the protein antigen and a biotinylated monoclonal antibody specific for target protein were simultaneously incubated. The streptavidin–peroxidase followed with a substrate solution was added after washing to induce a colored reaction. Colorimetric absorbance was read at 450 nm on a SpectraMax 190 Microplate reader (Molecular Devices, Sunnyvale, CA, USA). Serum CK18 concentration was expressed in U/L (1 U/L = 1.24 pM recombinant protein standard).

#### Test 2

CK18 levels were estimated by using in vitro immunoassay M30-Apoptosense ELISA kit (PEVIVA<sup>®</sup>, DiaPharma).

This one-step test is designed to measure soluble caspase-cleaved keratin 18 fragments in serum. K18Asp396 neo-epitope is termed M30 antigen. An antibody directly against it is used in this test.

One single researcher was responsible for all techniques, following manufacturer's instructions. Absorbance was read at 450 nm. Serum concentration was expressed in U/L; also, all values were according to standard curve ranges.

### Statistical Analysis

Data were analyzed using a statistical software program (IBM<sup>®</sup> SPSS<sup>®</sup> Statistics, version 22.0). Two-tailed Student's *t* test and ANOVA were calculated for pairwise comparison of continuous variable from both tests. Categorical variables were expressed in percentages and analyzed using Chi-square test or Fisher's exact test when applicable. Statistical significance was defined when  $p < 0.05$ , using two-tailed tests. Results from CK18 ELISA tests were compared using Pearson's correlation test or partial correlation Pearson tests controlling for the effect size. Advanced fibrosis and NASH prediction was estimated for each CK18 test using binary logistics regression analysis, and acquired confidential intervals were compared with those using bootstrap technique.  $R^2$ , variables' significance, and odds ratio were compared between models from both tests. The performance of each CK18 tests for prediction of NASH and advanced fibrosis was determined using area under receiver operating characteristic curve (AUROC).

## Results

### Patient Characteristics

The main demographic, clinical, and laboratory characteristics of the patients are presented in Table 1. Serum CK18 M30 measurements were taken using both Test 1 and Test 2 on a total of 172 subjects. Only Test 2 was performed on the serum of an additional 11 patients who were enrolled after Test 1 was performed, bringing the total number of patients with Test 2 done to 183. There was no difference in baseline characteristics between groups of patients who had both T1 and T2 performed versus those who only had T2 performed. NASH was diagnosed in 49 % of patients, according to criteria proposed by the NASH Clinical Research Network [25]. Patient age, ethnicity, and race were similar in patient with and without NASH. While 60 % of the total cohort was male, there were more women in NASH group than in the non-NASH group (47 vs. 34 %,  $p = 0.024$ ). Mean body mass index (BMI), ALT, and AST were higher in patients with NASH, otherwise diabetes and hypertension had similar frequencies.

**Table 1** Patient characteristics

	All patients ( <i>n</i> = 183)	NASH ( <i>n</i> = 90)	Non-NASH ( <i>n</i> = 93)	<i>p</i> value
Age (years)	50 ± 12	49 ± 12	51 ± 12	0.320
Sex (F/M) (%)	40/60	47/53	34/66	0.024
Ethnicity (H/NH) (%)	16/84	78/22	87/13	0.189
Race (W/B/A/O) (%)	87/2/2/9	84/1/1/14	89/3/4	0.470
BMI (kg/m <sup>2</sup> )	34 ± 7	35.4 ± 7.2	33.1 ± 6.1	0.023
Diabetes (%)	36.4	28.3	36.4	0.324
Hypertension (%)	52.4	52.1	52.4	0.454
ALT (IU/mL)	75.8 ± 50	86.5 ± 59.7	64.6 ± 35.8	0.004
AST (IU/mL)	50.6 ± 32	57.6 ± 40	42.8 ± 19.2	0.002
AST/ALT	0.72 ± 0.23	0.72 ± 0.24	0.72 ± 0.22	0.832
Albumin (g/dL)	4.5 ± 0.4	4.51 ± 0.44	4.53 ± 0.38	0.705
Bilirubin (mg/dL)	0.57 ± 0.039	0.55 ± 0.42	0.57 ± 0.29	0.634
Glucose (mg/dL)	117 ± 47	123 ± 52	112 ± 41	0.198

*H* Hispanic, *NH* non-Hispanic, *W* White, *B* Black, *A* Asian and *O* others

The histological features of subjects are presented in Table 2. Of the 49 % patients who had NASH diagnosed, 15.5 % had advanced fibrosis (grades 3 or 4); 31.6 % presented with severe inflammation (grade 2) and 24.6 % with severe ballooning (grade 2).

### CK18 Levels from Both Tests and Relation to Liver Histology Findings

Intra-assay variability was considered acceptable for both tests. CV% for Test 1 and Test 2 was 9.2 and 6.0, respectively.

While the mean serum CK18 level measured by Test 2 was significantly higher in patients with NASH, advanced fibrosis, or severe ballooning, the mean level measured by Test 1 was not significantly different between the different groups (see Table 3). CK18 level measured by Test 2 did not differ between low and severe inflammation, although a trend suggesting this difference was noticed ( $p = 0.053$ ).

One-way ANOVA and post hoc test were used to identify differences in CK18 levels among the various stages of steatosis, inflammation, ballooning, and fibrosis.

Significant differences were found just in CK18 from Test 2, except among inflammation stages. Test 1 was ineffective in distinguishing stages of any parameters.

There were no significant differences for Test 2 and no differences for Test 1 concerning inflammation stages. Steatosis: significant differences, using Test 2, between grade 1 versus 3 ( $p = 0.026$ ); ballooning: grade 0 versus 2 ( $p = 0.037$ ), 1 versus 2 ( $p = 0.049$ ); and fibrosis: grade 0 versus 2, 3, and 4 ( $p = 0.002, 0.005, 0.002$ ).

### Poor Inter-test Reliability Between CK18 Kits

There was no significant correlation between measurements from the two tests using the Pearson's correlation ( $p = 0.86, r = 0.01$ ) or the partial correlation Pearson tests controlling for the effect size of fibrosis, NAS score, and hepatocyte ballooning ( $p = 0.65, 0.89, \text{ and } 0.81$ , respectively).

Binary logistics regression was carried out to assess serum level of CK18 measured by both tests as a predictor of NASH and advanced fibrosis. The results are given in Table 4. Serum CK18 level from Test 2 was a significant

**Table 2** Histological findings

	All patients ( <i>n</i> = 183)	NASH ( <i>n</i> = 90)	Non-NASH ( <i>n</i> = 93)	<i>p</i> value
Specimen length (mm)	12.4 ± 6.5	13.5 ± 7.7	11.6 ± 4.6	0.044
Steatosis grade	1.9 ± 0.8	2.3 ± 0.6	1.5 ± 0.7	<0.001
Inflammation grade	1.4 ± 0.6	1.7 ± 0.6	1.0 ± 0.4	<0.001
Ballooning grade	0.9 ± 0.8	1.2 ± 0.9	0.5 ± 0.6	<0.001
Fibrosis grade	1.4 ± 1.3	1.8 ± 1.2	1.0 ± 1.1	<0.001
NAS score	4.5 ± 1.5	5.7 ± 0.8	3.4 ± 0.9	<0.001
Advanced fibrosis (%)	19.3	15.5	3.7	<0.001
Severe inflammation (%)	37.4	31.6	5.9	<0.001
Severe ballooning (%)	27.8	24.6	3.2	<0.001

**Table 3** CK18 concentration according to different liver histological parameters

	Test 1 (ng/L)			Test 2 (U/L)		
	Mean		<i>p</i> value	Mean		<i>p</i> value
	Yes	No		Yes	No	
NASH	613	442	0.31	460	285	0.001
Advanced fibrosis	610	488	0.5	529	304	0.001
Severe ballooning	728	382	0.07	456	316	0.02
Advanced inflammation	1476	823	0.33	422	298	0.053

predictor of NASH and advanced fibrosis, whereas CK18 level from Test 1 was not.

To assess the role of CK18 as a diagnostic tool for the prediction of NASH and advanced fibrosis, the receiver operating characteristic curves (ROCs) were constructed and are shown in Fig. 1. Again, Test 1 performed very poorly for both the prediction of NASH and advanced fibrosis with the area under the ROC (AUROC) of 0.513 (95 % CI 0.425–0.601) for NASH and 0.517 (95 % CI 0.408–0.627) for advanced fibrosis. Test 2 performed better, with AUROC of 0.638 (95 % CI 0.555–0.722) for NASH and 0.676 (95 % CI 0.571–0.782) for advanced fibrosis.

Table 5 lists the overall sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios of CK18 by Test 2 for NASH and advanced fibrosis. All these values were calculated using the optimal cutoff points of CK18 levels after Youden’s index analysis (356.18 and 395.97 U/L for NASH and advanced fibrosis diagnosis, respectively). Test 2 has poor sensitivity for both NASH and advanced fibrosis, showing that CK18 as a biomarker alone will miss half of the NASH cases.

**Discussion**

The aim of this study was to assess the inter-test reliability between two CK18 ELISA kits in a cohort of patients with biopsy-proven NAFLD and to evaluate the test performance of this biomarker in the prediction of NASH and advanced fibrosis in patients with NAFLD. This study

highlights the absence of correlation between the results from two commercially available CK18 kits.

Given the high prevalence of NAFLD and the variable severity and disease course, a simple, quick inexpensive biomarker for risk stratification and to monitor progression is very appealing. Given the previous data suggesting CK18 as a practical and reproducible noninvasive biomarker in NASH [17, 22, 26], it is increasingly used in clinical trials. Our study confirms that serum levels of CK18 fragments are significantly higher in NASH and advanced fibrosis ( $p = 0.001$ ,  $p = 0.001$ , respectively) when Test 2 is used. However, our results showed that Test 1 was essentially useless as a biomarker in NAFLD.

Although statistical differences were demonstrated between these groups, low sensitivity (Test 1 50 % and Test 2 60 %) and specificity (Test 1 79 % and Test 2 75 %) showed that this biomarker alone was insufficiently accurate as a predictor of NASH and advanced fibrosis. Previous studies, using the same Test 2 for the diagnosis of NASH, achieved sensitivity of 56–77 % and specificity of 63–92 %, demonstrating a wide range despite using the same kit [22–24]. Using the optimal serum CK18 cutoffs for our cohort (356.18 U/L for NASH and 395.97 U/L for advanced fibrosis diagnosis), the accuracy of Test 2 is 64.6 and 71.8 %, respectively, resulting in the misclassification of 35.4 and 28.2 % patients for NASH and advanced fibrosis, respectively.

The AUROC was estimated to assess the role of CK18 M30 as a diagnostic tool for the prediction of NASH and advanced fibrosis (0.64 and 0.67, respectively). Although the multicenter validation study published by Feldstein et al. showed higher AUROC for prediction of NASH (0.83), most of recent studies demonstrated lower values ranging from 0.53 to 0.63 for NASH and 0.53 to 0.68 for fibrosis prediction [23, 24, 27]. One possible explanation for this contradictory result is the different percentages of subjects with NASH among studies—49 % of our cohort had NASH, compared to just 19 % of the previous validation cohort. All of the studies presented above used the same kit, suggesting that besides variability between kits, other factors such as differences in the study cohorts and test reliability may be at play.

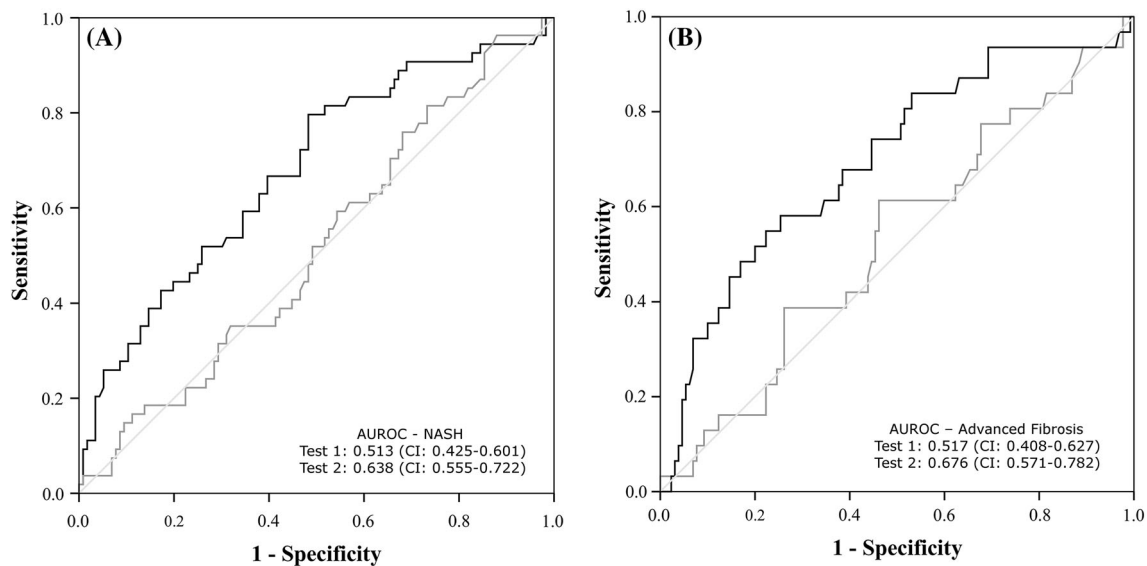
Recent studies [23, 24] report poor performance of CK18 fragments as a biomarker to identify NASH and

**Table 4** Logistics regression model for prediction of different liver histology parameters according to Test 1 and Test 2

Prediction	Test 1			Test 2		
	$b^a$	<i>p</i>	OR (CI)	$b^a$	<i>p</i>	OR (CI)
NASH	0.0001	0.31	1.0 (1.0–1.0)	0.001	0.005	1.001 (1–1.002)
Advanced fibrosis	0.00009	0.56	1.0 (1.0–1.0)	0.001	0.004	1.001 (1–1.002)

<sup>a</sup> Value of the model equation coefficient





**Fig. 1** **a** A comparison of CK18 by Test 1 (gray line) and by Test 2 (black line) in predicting NASH. **b** A comparison of CK18 by Test 1 (gray line) and by Test 2 (black line) in predicting advanced fibrosis

**Table 5** CK18 by Test 2 performance for the diagnosis of NASH and advanced fibrosis

	NASH	Advanced fibrosis
Sensitivity	50 %	60 %
Specificity	79 %	75 %
Positive predictive value	68 %	36 %
Negative predictive value	62 %	89 %
Likelihood ratio (+)	2.4	2.4
Likelihood ratio (−)	0.6	0.5

advanced fibrosis in NAFLD patients supported by low AUROC, sometimes similarly to conventional biochemistry evaluation like AST and ALT. Chan et al. [23] highlight the limited utility of M30 in the detection of NASH based on low AUROC for NASH (0.59), lower than the AUROC achieved by ALT (0.64) and AST (0.75) in the same cohort of patients. Similarly, Cusi et al. [24] concluded that although CK18 M30 has a reasonable sensitivity for NAFLD and any stage of fibrosis (68 and 85 %), its low specificity (58 and 54 %, respectively) rendered it a poor test for screening and staging NASH. Also, the same study reinforced the low AUROC for prediction of NASH (0.65) and the presence of fibrosis (0.68).

The discordance between the different studies of CK18 test performance as a biomarker of NASH diagnosis may be influenced by different ethnic population, size of cohorts, prevalence of diabetes, and other comorbidities and unbalance between forms of disease presentation (simple steatosis, NAFLD, and NASH). All these factors

(stages 3–4). The predicted AUROC with its 95 % CI was shown below each panel

together prevent a clear interpretation about the discordance among studies.

In our study, the noticeable lack of correlation between the two tests in the same cohort highlights the importance of standardization of all the available kits and may explain the differences in findings between different studies. However, given the wide variability between two different kits in the same cohort shown by our study, we suggest that the kit's influence is significant and should not be undervalued.

CK18 kits are only available as a research tool and not yet evaluated for clinical purpose in the setting of NASH disease. Thus, this method did not undergo through the extensive protocols involved during regulatory entities validation and approval. Further analysis is needed to determine batch-to-batch variability and variability with repeated freeze/thaw cycles. This is particularly important in large studies that work with large cohorts through long periods of follow-up, where tests will also be run in multiple batches.

Further studies on the standardization of CK18 kits are needed to determine which kit is most reliable and has the best test performance as a biomarker in NAFLD. In addition, combinations of several serologic markers with better performance to identify NASH or advanced fibrosis are a promising alternative [20, 21]. While serum CK18 fragments is insufficiently accurate as a biomarker in NAFLD, its measurement using the right ELISA kit remains promising as an important tool for NAFLD evaluation.

This study does have some limitations. The average biopsy length ( $12.4 \pm 6.5$  cm) might be considered low

according to guidelines parameters [9, 25], but also as the main end point of our study is to evaluate CK18 M30 test performance; thus, correlations between biomarker levels were not affected by histological features. Our study used histopathological examination of liver biopsies as a reference to stages liver fibrosis, which may underestimate severity due to sample error and intra-observer variability. Another weakness of this study is not have a control group; comparisons were made between NASH versus non-NASH groups and patients with and without advanced features in liver biopsy.

This study is the first study comparing serum CK18 fragment levels between different kits in a large group of well-characterized NAFLD patients with biopsy-proven NASH. In summary, our findings show that while serum CK18 M30 level using the right test kit has a role in the noninvasive assessment of NAFLD, there are significant variations between ELISA kits, which could greatly bias the results. There is a need for standardization of the multiple available CK18 ELISA kits.

**Acknowledgments** This publication was supported by NIDDK Grant 1K23DK083439 to Michelle Lai, MD (PI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDDK or the NIH.

**Author contributions** C.F.M.G.P. and M.L. were involved in the primary study design and data analysis; C.F.M.G.P., Z.G.J., T.O., and L.F. were involved in CK18 level measurements; T.L.C. and I.N. were responsible for liver pathology analysis; S.R. and N.A. were responsible for manuscript analysis and comments; all authors participated in the preparation of this manuscript.

**Compliance with ethical standards**

**Conflict of interest** None.

## References

- Vernon G, Baranova A, Younossi ZM. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. *Aliment Pharmacol Ther.* 2011;34:274–285.
- Browning JD, Szczepaniak LS, Dobbins R, et al. Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology.* 2004;40:1387–1395.
- Williams CD, Stengel J, Asike MI, et al. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. *Gastroenterology.* 2011;140:124–131.
- Younossi ZM, Stepanova M, Afendy M, et al. Changes in the prevalence of the most common causes of chronic liver diseases in the United States From 1988 to 2008. *Clin Gastroenterol Hepatol.* 2011;9:524–530.
- Adams LA, Sanderson S, Lindor KD, Angulo P. The histological course of nonalcoholic fatty liver disease: a longitudinal study of 103 patients with sequential liver biopsies. *J Hepatol.* 2005;42:132–138.
- Teli MR, James OF, Burt AD, Bennett MK, Day CP. The natural history of nonalcoholic fatty liver: a follow-up study. *Hepatology.* 1995;22:1714–1719.
- Starley BQ, Calcagno CJ, Harrison SA. Nonalcoholic fatty liver disease and hepatocellular carcinoma: a weighty connection. *Hepatology.* 2010;51:1820–1832.
- Bugianesi E, Leone N, Vanni E, et al. Expanding the natural history of nonalcoholic steatohepatitis: from cryptogenic cirrhosis to hepatocellular carcinoma. *Gastroenterology.* 2002;123:134–140.
- Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Gastroenterological Association, American Association for the Study of Liver Diseases, and American College of Gastroenterology. *Gastroenterology.* 2012;142:1592–1609.
- Ong JP, Younossi ZM. Epidemiology and natural history of NAFLD and NASH. *Clin Liver Dis.* 2007;11:1–16.
- Singh S, Allen AM, Wang Z, Prokop LJ, Murad MH, Loomba R. Fibrosis progression in nonalcoholic fatty liver vs nonalcoholic steatohepatitis: a systematic review and meta-analysis of paired-biopsy studies. *Clin Gastroenterol Hepatol.* 2015;13:643.e9–654.e9.
- McPherson S, Hardy T, Henderson E, Burt AD, Day CP, Anstee QM. Evidence of NAFLD progression from steatosis to fibrosis-steatohepatitis using paired biopsies: implications for prognosis and clinical management. *J Hepatol.* 2014;62:1148–1155.
- Pais R, Charlotte F, Fedchuk L, et al. A systematic review of follow-up biopsies reveals disease progression in patients with non-alcoholic fatty liver. *J Hepatol.* 2013;59:550–556.
- 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;61:1547–1554.
- Kramer G, Erdal H, Mertens HJMM, et al. Differentiation between cell death modes using measurements of different soluble forms of extracellular cyokeratin 18. *Cancer Res.* 2004;64:1751–1756.
- Wieckowska A, Zein NN, Yerian LM, Lopez AR, McCullough AJ, Feldstein AE. In vivo assessment of liver cell apoptosis as a novel biomarker of disease severity in nonalcoholic fatty liver disease. *Hepatology.* 2006;44:27–33.
- Yilmaz Y. Systematic review: caspase-cleaved fragments of cyokeratin 18—the promises and challenges of a biomarker for chronic liver disease. *Aliment Pharmacol Ther.* 2009;30:1103–1109.
- Aida Y, Abe H, Tomita Y, et al. Serum cyokeratin 18 fragment level as a noninvasive biomarker for non-alcoholic fatty liver disease. *Int J Clin Exp Med.* 2014;7:4191–4198.
- Younossi ZM, Jarrar M, Nugent C, et al. A novel diagnostic biomarker panel for obesity-related nonalcoholic steatohepatitis (NASH). *Obes Surg.* 2008;18:1430–1437.
- Cao W, Zhao C, Shen C, Wang Y. Cyokeratin 18, alanine aminotransferase, platelets and triglycerides predict the presence of nonalcoholic steatohepatitis. *PLoS One.* 2013;8:1–8.
- Anty R, Iannelli A, Patouraux S, Bonnafous S, Lavallard VJ, Amor IB. A new composite model including metabolic syndrome, alanine aminotransferase and cyokeratin-18 for the diagnosis of non-alcoholic steatohepatitis in morbidly obese patients. *Aliment Pharmacol Ther.* 2010;32:1315–1322.
- Feldstein AE, Wieckowska A, Lopez AR, Liu YC, Zein NN, McCullough AJ. Cyokeratin-18 fragment levels as noninvasive biomarkers for nonalcoholic steatohepatitis: a multicenter validation study. *Hepatology.* 2009;50:1072–1078.
- Chan W-K, Sthaneshwar P, Mustapha NRN, Mahadeva S. Limited utility of plasma M30 in discriminating non-alcoholic steatohepatitis from steatosis—a comparison with routine biochemical markers. *PLoS One.* 2014;9:e105903.
- Cusi K, Chang Z, Harrison S, et al. Limited value of plasma cyokeratin-18 as a biomarker for NASH and fibrosis in patients with non-alcoholic fatty liver disease. *J Hepatol.* 2014;60:167–174.

25. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*. 2005;41:1313–1321.
26. Yilmaz Y, Dolar E, Ulukaya E, et al. Soluble forms of extracellular cytokeratin 18 may differentiate simple steatosis from nonalcoholic steatohepatitis. *World J Gastroenterol*. 2007; 13:837–844.
27. Shen J, Chan HL-Y, Wong GL-H, et al. Assessment of non-alcoholic fatty liver disease using serum total cell death and apoptosis markers. *Aliment Pharmacol Ther*. 2012;36:1057–1066.