**Name of Journal:** *World Journal of Gastroenterology*

**Manuscript NO:** 72131

**Manuscript Type:** LETTER TO THE EDITOR

**Comments on validation of conventional non-invasive fibrosis scoring systems in patients with metabolic associated fatty liver disease**

Hong JG *et al*. MAFLD and non-invasive fibrosis scoring systems

**Jian-Guo Hong, Lun-Jie Yan, Xian Li, Sheng-Yu Yao, Peng Su, Hai-Chao Li, Zi-Niu Ding, Dong-Xu Wang, Zhao-Ru Dong, Tao Li**

**Jian-Guo Hong, Lun-Jie Yan, Xian Li, Sheng-Yu Yao,** **Hai-Chao Li, Zi-Niu Ding, Dong-Xu Wang, Zhao-Ru Dong, Tao Li,**Department of General Surgery, Qilu Hospital of Shandong University, Jinan 250012, Shandong Province, China

**Peng Su,**Department of Pathology, Qilu Hospital of Shandong University, Jinan 250012, Shandong Province, China

**Author contributions:** Li X and Yao SY designed and performed the research; Yan LJ, LiHC and Ding ZN analyzed the data; Li T and Hong JG wrote the letter; Yan LJ, Su P, Wang DX and Dong ZR revised the letter; Hong JG, Yan LJ, Li X, Yao SY, and Su P contributed equally to this article.

**Corresponding author: Tao Li, MD,** **Chief Doctor,** Department of General Surgery, Qilu Hospital of Shandong University, No. 107 West Wenhua Road, Jinan 250012, Shandong Province, China. litao7709@126.com

**Received:** October 5, 2021

**Revised:** December 5, 2021

**Accepted: January 17, 2022**

**Published online:**

**Abstract**

To evaluate and predict liver fibrosis in patients with nonalcoholic fatty liver disease (NAFLD), several non-invasive scoring systems were built and widely used in the progress of diagnosis and treatment, which showed great diagnostic efficiency, such as aspartate aminotransferase to platelet ratio index, fibrosis-4 index, body mass index, aspartate aminotransferase to alanine aminotransferase ratio, diabetes score and NAFLD fibrosis score. Since the new concept of metabolic associated fatty liver disease (MAFLD) was proposed, the clinical application value of the non-invasive scoring systems mentioned above has not been assessed in MAFLD. The evaluation of the diagnostic performance of these non-invasive scoring systems will provide references for clinicians in the diagnosis of MAFLD.

**Key Words:** Metabolic associated fatty liver disease; Prediction model; Calibration; Normal distribution; Nonalcoholic fatty liver disease

Hong JG, Yan LJ, Li X, Yao SY, Su P, Li HC, Ding ZN, Wang DX, Dong ZR, Li T. Comments on validation of conventional non-invasive fibrosis scoring systems in patients with metabolic associated fatty liver disease. *World J Gastroenterol* 2022; In press

**Core Tip:** The concept of metabolic associated fatty liver disease (MAFLD) was proposed in 2020. Unlike the concept of nonalcoholic fatty liver disease, the exclusion of chronic liver disease was not required in the establishment of diagnosis of MAFLD, but the presence of metabolic associated disease or dysfunction is required. The clinical prediction values and the optimal cutoff values of non-invasive fibrosis scores remain unknown. We read the recent article entitled “Validation of Conventional Non-invasive Fibrosis Scoring Systems in Patients with Metabolic Associated Fatty Liver Disease” with great interest. We would like to share our opinions and criticisms about this valuable work.

**TO THE EDITOR**

We read the recent article entitled “Validation of Conventional Non-invasive Fibrosis Scoring Systems in Patients with Metabolic Associated Fatty Liver Disease*”* published by Wu *et al*[1] with great interest. In the article, the authors designed and performed a retrospective study to evaluate the diagnostic performance of four non-invasive scoring systems, including aspartate aminotransferase to platelet ratio index (APRI), fibrosis-4 index (FIB-4), body mass index (BMI), aspartate aminotransferase to alanine aminotransferase (ALT) ratio, diabetes score (BARD score) and nonalcoholic fatty liver disease fibrosis score (NFS), in patients with metabolic associated fatty liver disease (MAFLD). We would like to share our opinions and criticisms about this valuable work.

The specificity, sensitivity, positive predictive value (PPV), negative predictive value (NPV) and discrimination of the scoring systems mentioned above were evaluated in the prediction of advanced fibrosis in patients with MAFLD in the study by Wu *et al*[1]. Clinical characteristics, laboratory variables and non-invasive scores were compared between patients with advanced fibrosis and those with mild fibrosis or without fibrosis. The results showed that the FIB-4 (*P* < 0.001), NFS (*P* < 0.001), APRI (*P* = 0.003) and BARD (*P* < 0.001) scores were all significantly higher in patients with advanced fibrosis. Unfortunately, only univariate analysis was performed in this study. In our opinion, multivariate analysis should be performed to recognize the independent variables in the prediction of advanced fibrosis, as in the study by Nielsen *et al*[2].

The authors evaluated the diagnostic efficiency of the prediction scores using the following statistical indices: Sensitivity, specificity, accuracy, PPV, NPV and the area under the receiver operating characteristic curve (AUROC). All the above indices are important statistical variables in the development and validation of prediction models. In our opinion, it would be better if the authors had evaluated the calibration of the prediction scores in the study. In fact, calibration of the prediction model is a critical statistical index in the evaluation of diagnostic efficiency[3]. Calibration of the prediction scores can be performed using the Hosmer–Lemeshow goodness-of-fit test and calibration curves, the latter of which can be easily plotted using R software. We advise the authors to evaluate the calibration of prediction models in future studies.

The calculations of PPV and NPV are very important in the development and validation of prediction models. Unlike sensitivity and specificity, PPV and NPV cannot be compared directly among different samples, except for samples with the same prevalence rate of the disease. This is because both PPV and NPV can be affected by the prevalence rate of disease[4,5]. The authors compared the PPV and NPV in table 4 and stated that “PPV and NPV was better in the HBV-MAFLD group” in the article. In our opinion, the comparisons of PPV and NPV between the HBV-MAFLD group and the pure MAFLD group will be valuable only when advanced fibrosis accounts for the same proportion of the two groups.

Although the authors stated that the continuous variables were expressed in the format of mean ± SD or median value with interquartile range (IQR) and the differences were calculated using Student’s *t* test in the case of normally distributed data or the Mann–Whitney test in the remaining cases, there were no continuous variables expressed as the median (IQR) in the article. Normally, the distribution of continuous variables should be tested; then, the continuous variables in normal distribution will be expressed as the mean ± SD, and the non-normally distributed continuous variables will be expressed as the median (IQR). If all the continuous variables are expressed as the mean ± SD but the authors do not indicate that all the continuous variables fit a normal distribution, the readers will doubt whether the normal distribution tests were performed in the study. After all, it rarely happens that all the laboratory variables fit a normal distribution in one study. In most studies, the laboratory variables and scores, including ALT, AST, APRI and other variables, do not fit a normal distribution and should be expressed as the median (interquartile range)[2,6-8]. For any parameter in biomedical research, a true normal distribution is rare. The European Medicines Agency has issued the general guidance that data should be checked for normality of distribution and should be analyzed and presented based on the results of normal distribution tests. It is also possible that some continuous variables did not fit a normal distribution, but these variables were accidentally expressed as the mean ± SD in the study conducted by Wu *et al*[1]. Of course, this situation indeed does not affect the accuracy of the study. We advise that the authors indicate whether the continuous variables fit a normal distribution and if the normal distribution tests have been performed in future studies.

The authors compared the diagnostic ability of the NFS, FIB-4, APRI and BARD score for a late stage of fibrosis in MAFLD. The results demonstrated that the APRI and BARD scores performed poorly, but the FIB-4 and NFS showed a promising prospect in clinical use. The new thresholds of the FIB-4 and NFS proposed in this study were 1.05 and -2.1, respectively. The two thresholds proposed by the authors were determined based on their specific study sample. The diagnostic efficiency of the thresholds in the prediction of advanced fibrosis should be further evaluated in an external validation cohort and/or in a prospective validation cohort. Additionally, if possible, the authors can try to develop models based on multiple variables, including the FIB-4 and/or NFS, to predict advanced fibrosis in patients with MAFLD. He *et al*[9] proposed that a diagnostic model containing valuable parameters extracted from more examination tools might provide more satisfactory results[9]. Compared to using a single variable, we believe that prediction models based on multiple variables, including clinical characteristics, radiology examinations and laboratory examinations, would exhibit higher sensitivity, higher specificity, higher accuracy, higher PPV, higher NPV, better discrimination and better calibration in the prediction of advanced fibrosis in patients with MAFLD. Because there is now evidence from a prospective cohort that common genetic variants can capture additional prognostic insights not conveyed by validated clinical/biochemical parameters[10], we encourage the integration of genetics (perhaps epigenetics) with clinical fibrosis scores, as it may refine individual risk and improve risk stratification and prediction of severe liver disease.

In general, we are very interested in the study by Wu *et al*[1]. The authors demonstrated the prediction values of APRI, FIB-4, NFS and BARD in a large sample of histology-proven MAFLD. As MAFLD is a new entity, this study will provide important references for clinicians in the prediction of advanced fibrosis in MAFLD patients. The study performed by Wu *et al*[1] could also provide important references for other studies of non-invasive scores and prediction models.

**REFERENCES**

1 **Wu YL**, Kumar R, Wang MF, Singh M, Huang JF, Zhu YY, Lin S. Validation of conventional non-invasive fibrosis scoring systems in patients with metabolic associated fatty liver disease. *World J Gastroenterol* 2021; **27**: 5753-5763 [PMID: 34629799 DOI: 10.3748/wjg.v27.i34.5753]

2 **Nielsen MJ**, Leeming DJ, Goodman Z, Friedman S, Frederiksen P, Rasmussen DGK, Vig P, Seyedkazemi S, Fischer L, Torstenson R, Karsdal MA, Lefebvre E, Sanyal AJ, Ratziu V. Comparison of ADAPT, FIB-4 and APRI as non-invasive predictors of liver fibrosis and NASH within the CENTAUR screening population. *J Hepatol* 2021; **75**: 1292-1300 [PMID: 34454994 DOI: 10.1016/j.jhep.2021.08.016]

3 **Assel M**, Sjoberg D, Elders A, Wang X, Huo D, Botchway A, Delfino K, Fan Y, Zhao Z, Koyama T, Hollenbeck B, Qin R, Zahnd W, Zabor EC, Kattan MW, Vickers AJ. Guidelines for Reporting of Statistics for Clinical Research in Urology. *Eur Urol* 2019; **75**: 358-367 [PMID: 30580902 DOI: 10.1016/j.eururo.2018.12.014]

4 **Monaghan TF**, Rahman SN, Agudelo CW, Wein AJ, Lazar JM, Everaert K, Dmochowski RR. Foundational Statistical Principles in Medical Research: Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value. *Medicina (Kaunas)* 2021; **57** [PMID: 34065637 DOI: 10.3390/medicina57050503]

5 **Akobeng AK**. Understanding diagnostic tests 2: likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta Paediatr* 2007; **96**: 487-491 [PMID: 17306009 DOI: 10.1111/j.1651-2227.2006.00179.x]

6 **Abdel-Hameed EA**, Rouster SD, Kottilil S, Sherman KE. The Enhanced Liver Fibrosis Index Predicts Hepatic Fibrosis Superior to FIB4 and APRI in HIV/HCV Infected Patients. *Clin Infect Dis* 2021; **73**: 450-459 [PMID: 32459305 DOI: 10.1093/cid/ciaa646]

7 **Lee J**, Vali Y, Boursier J, Spijker R, Anstee QM, Bossuyt PM, Zafarmand MH. Prognostic accuracy of FIB-4, NAFLD fibrosis score and APRI for NAFLD-related events: A systematic review. *Liver Int* 2021; **41**: 261-270 [PMID: 32946642 DOI: 10.1111/liv.14669]

8 **Amiri M**, Murgas S, Stang A, Michel MC. Do overactive bladder symptoms and their treatment-associated changes exhibit a normal distribution? Implications for analysis and reporting. *Neurourol Urodyn* 2020; **39**: 754-761 [PMID: 31945207 DOI: 10.1002/nau.24275]

9 **He Y**, Zhu Z, Chen Y, Chen F, Wang Y, Ouyang C, Yang H, Huang M, Zhuang X, Mao R, Ben-Horin S, Wu X, Ouyang Q, Qian J, Lu N, Hu P, Chen M. Development and Validation of a Novel Diagnostic Nomogram to Differentiate Between Intestinal Tuberculosis and Crohn's Disease: A 6-year Prospective Multicenter Study. *Am J Gastroenterol* 2019; **114**: 490-499 [PMID: 30741735 DOI: 10.14309/ajg.0000000000000064]

10 **De Vincentis A**, Tavaglione F, Jamialahmadi O, Picardi A, Antonelli Incalzi R, Valenti L, Romeo S, Vespasiani-Gentilucci U. A Polygenic Risk Score to Refine Risk Stratification and Prediction for Severe Liver Disease by Clinical Fibrosis Scores. *Clin Gastroenterol Hepatol* 2021 [PMID: 34091049 DOI: 10.1016/j.cgh.2021.05.056]

**Footnotes**

**Conflict-of-interest statement:** The authors have no conflicts of interest to disclose.

**Open-Access:** This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: https://creativecommons.org/Licenses/by-nc/4.0/

**Provenance and peer review:** Invited article; Externally peer reviewed.

**Peer-review model:** Single blind

**Peer-review started:** October 5, 2021

**First decision:** November 15, 2021

**Article in press:**

**Specialty type:** Gastroenterology and hepatology

**Country/Territory of origin:** China

**Peer-review report’s scientific quality classification**

Grade A (Excellent): A

Grade B (Very good): B

Grade C (Good): 0

Grade D (Fair): 0

Grade E (Poor): 0

**P-Reviewer:** Par G, Zaiou M **S-Editor:** Fan JR **L-Editor:** A **P-Editor:** Fan JR