

External Validation and Clinical Utility of Non-Invasive Prediction Models for Non-Alcoholic Fatty Liver Disease in Malaysia

Mohd Azri Mohd Suan (✉ irzah96@yahoo.com)

Hospital Sultanah Bahiyah <https://orcid.org/0000-0002-9508-7297>

Huan Keat Chan

Hospital Sultanah Bahiyah

Shahrul Aiman Soelar

Hospital Sultanah Bahiyah

Muhammad Radzi Abu Hassan

Hospital Sultanah Bahiyah

Research

Keywords: calibration, elastography, non-alcoholic fatty liver disease, predictive value, steatosis

Posted Date: November 6th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-102663/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Many prediction models have been developed to detect non-alcoholic fatty liver disease (NAFLD). The drawbacks of many of models are the use of parameters that are not routinely measured locally. This study aimed to evaluate the external validity of a series of prediction models for NAFLD, which were selected based on the routinely measured and tested clinical parameters in public healthcare centers in Malaysia.

Methods: A literature search of articles that described the prediction models for NAFLD on adult subjects between 2000 and 2019 was conducted. The validation cohort comprised patients who underwent liver elastography using the Fibroscan® device in a public tertiary care center between January 2017 and December 2019. Both the discrimination and calibration of each model were assessed to determine their predictive performance.

Results: Out of the 404 patients undergoing liver elastography, 280 were diagnosed with NAFLD (69.3%). Six prediction models were identified from the existing literature and evaluated. The calibration assessment demonstrated that although three of the models overestimated the NAFLD risk, updating the models generally improved their calibration performance. The discriminative performance of the selected models ranged from 0.717 to 0.783. With a specificity level of 90% and 80%, the sensitivity of all the models fell between 31.1%–48.9% and 46.4%–66.8%, respectively. The Framingham Steatosis Index (FSI) model demonstrated a better predictive performance compared to the other models.

Conclusions: The FSI model demonstrates an acceptable predictive performance. Its application in clinical practice could promote the screening and early treatment of NAFLD in the Malaysian population.

Background

Non-alcoholic fatty liver disease (NAFLD) refers to the presence of hepatic steatosis without a history of excessive alcohol consumption [1]. It comprises a broad clinicopathological spectrum from non-alcoholic fatty liver (NAFL) to non-alcoholic steatohepatitis (NASH). NAFLD has become an increasing public health concern in both developed and developing countries [2-5]. Approximately a quarter of the Asian population are currently living with NAFLD, which is mainly due to their sedentary lifestyle and unhealthy dietary habits [2]. In addition, NAFLD has emerged as the most common cause of fatal liver diseases, such as cryptogenic cirrhosis and hepatocellular carcinoma (HCC) [6]. It is projected to take over viral hepatitis as the major risk factor of HCC in the near future due to its increasing incidence [7].

Early-stage NAFLD is usually asymptomatic, which makes it difficult to diagnose. The gold standard for an NAFLD diagnosis is a histologic assessment using a liver biopsy. However, this invasive procedure has several limitations, including its requirement for adequate tissue samples [8], limited accessibility [9], and the risk of complications following the procedure [10]. Hence, the diagnosis of NAFLD via the ultrasonography or transient elastography is becoming more common.

To address the limitations of the existing measures, many non-invasive prediction models have been developed to guide the risk assessment of NAFLD. The parameters used in these models range from demographic characteristics, anthropometry measurements, and laboratory findings to more specific biomarkers, such as sphingolipids and sterols. The drawbacks of many of these models are the use of parameters that are not routinely measured and tested and the high costs that are involved, which limits their feasibility in healthcare settings.

In Malaysia, the prevalence of NAFLD is high, estimated to fall between 37.4% and 46.0% [11-13]. The disease is more common in those who have diabetes (49.6%) [14], hypercholesterolaemia (56.7%) [15], and metabolic syndrome (82.8%) [16]. However, a national screening program for NAFLD is still not in place. The uncertainty about the usefulness of the screening tools and the high cost of the screening tests commonly preclude NAFLD screening in the clinical practice [17]. Furthermore, imaging tests for the NAFLD confirmatory diagnosis were only available in tertiary care centers, which generally have a high patient load and a long waiting time. These limitations often result in missed opportunities to detect early-stage NAFLD, which can be reversed with improved dietary habits and lifestyle changes. Therefore, a reliable and handy prediction model for NAFLD would be beneficial. This study aims to validate a range of prediction models for NAFLD, particularly those that apply the routinely measured and tested parameters in public healthcare centers in Malaysia.

Methods

This study received approval from the Medical Research and Ethics Committee under the Ministry of Health Malaysia (NMRR-20–748-54587). The requirement for informed consent was waived because the data were retrospectively accessed. The study's methods and findings were in line with the guidelines on the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) [18].

Selection of prediction models

A literature search was performed in February 2020 to identify the prediction models that were developed for NAFLD from the PubMed. The search strategy used the following search string: (fatty liver [Title/Abstract]) OR (NAFLD [Title/Abstract]) OR (steatosis [Title/Abstract]) AND (predict [Title/Abstract]) OR (index [Title/Abstract]) OR (risk [Title/Abstract]) OR (score [Title/Abstract]) OR (model [Title/Abstract]) OR (algorithm [Title/Abstract]) OR (test [Title/Abstract]) OR (biomarker [Title/Abstract]) OR (machine learning [Title/Abstract]). The search was limited to full-text articles regarding research on adult subjects that were written in English and published between 2000 and 2019. An article was only selected if it did the following: (i) presented the development of a prediction model or an update of a previously developed model for NAFLD, (ii) used the risk of developing NAFLD in the general population as the study's endpoint, (iii) applied multiple parameters or predictors to the model, (iv) developed the model based on weighted risk predictors, (v) provided the full model's linear predictor or prediction algorithm, and (vi) only applied parameters that were routinely measured and tested in public healthcare centers in Malaysia. When a full-

text article was not made publicly available, we made up to two attempts to approach the corresponding authors by email. The reference lists of the selected articles were also used to identify additional relevant articles.

Validation cohort

The validation cohort comprised patients seeking care from Hospital Sultanah Bahiyah, a public tertiary care center, which also served as the gastroenterology referral center in northern Malaysia. They were all above 18 years of age and underwent liver elastography using the Fibroscan® device (EchoSens, Paris) between January 2017 and December 2019. Those who had a history of active alcohol consumption (more than 14 drinks per week for men or more than seven for women), viral hepatitis, autoimmune hepatitis, or other forms of chronic liver disease were excluded.

The information on the risk factors or predictors that were used in each prediction model was obtained from the patient's electronic medical records. The predictors included individual socio-demographic and clinical information, ranging from age, ethnicity, gender, education level, marital status, occupation, and body mass index (BMI) to the presence of cardiovascular diseases (diabetes mellitus, hypertension, dyslipidaemia, and coronary artery disease) and the laboratory findings, including alanine aminotransferase (ALT), aspartate aminotransferase (AST), fasting blood glucose level, triglycerides (TG), and serum cholesterol.

The diagnosis of NAFLD was confirmed by physicians based on the findings of the liver elastography. The controlled attenuation parameter (CAP) was used to measure the level of hepatic steatosis—a reading above 248 decibels/meter (dB/m) indicated NAFLD [19]. Ten measurements were performed for each patient, and the diagnosis was only confirmed if at least six readings were valid. For the purpose of this study, only the CAP results from the first liver elastography, as well as the information and laboratory findings from the patient's clinic visits prior to the first liver elastography, were gathered.

Statistical analysis

Generally, studies on the external validation of prediction models require at least 100 (or ideally more than 200) events to generate an adequate study sample size [20,21]. To make up for the incomplete information in the validation cohort, the predictive mean matching method was applied to generate five imputed datasets, which were then pooled using Rubin's rules [22]. The demographic and clinical characteristics of the patients were summarized as either percentages (categorical data) or means and standard deviations (numerical data).

For each patient in the validation cohort, their risk of NAFLD was calculated using the algorithms provided by the selected prediction models. The predictive performance of each model was estimated using discrimination (the ability of a model to differentiate between individuals with and without NAFLD) and calibration (the agreement between the predictions and observed outcomes). The model's calibration was assessed graphically using a calibration plot. A perfect model prediction was expected to be

represented by a 45° line with an intercept (α) of zero and a slope (β) of one in the calibration plot [23]. The calibration intercept quantified the degree of agreement between the proportion of observed NAFLD cases and the mean predicted probability, which would indicate whether the predictions were systematically too low or too high [23]. On the other hand, the calibration slope referred to the degree of agreement between the predicted probability of developing NAFLD in the present study and the actual probability of having NAFLD [24]. The graph for each model was plotted based on the results of ten groups of a similar number of patients from the validation cohort who had similar predicted probabilities [24].

Direct application of the published models on the current validation cohort might have caused miscalibration, which is characterized by deviations from the ideal line (i.e. calibration-in-the-large was not equal to zero and the calibration slope was less or more than one). In the case of model miscalibration, the prediction model was updated by calculating a correction factor using the following equation [25]:

The correction factor was then added to the original model's intercept, and the new intercept was used when the updated model was applied to the validation cohort.²⁵ This method improved the model's calibration without affecting its ability to discriminate between individuals with and without NAFLD [26]. Furthermore, the model's discrimination was assessed based on the concordance ('c') statistic, which was equal to the area under the receiver operating characteristic (ROC) curve, along with its corresponding 95% confidence interval. Areas under the ROC that were greater than 0.5 suggested that the model could be used to predict NAFLD [27].

Subsequently, the diagnostic accuracy for each updated prediction model was examined using the sensitivity, specificity, positive- and negative-likelihood ratios and the positive and negative predictive values. These diagnostic parameters were calculated using a cut-off value that meant that ten percent of the population had values above the model's cut-off points. The procedure was then repeated using cut-offs where 20%, 80%, and 90% had values above the cut-off. All the data in this study was analyzed using the R statistical software version 3.5.2 (rms, Hmisc, pROC and rmda packages) [28].

Results

Prediction model selection

The search yielded 5985 articles from PubMed, 5899 of which were excluded based on their titles or abstracts. A total of 86 articles fulfilled the inclusion criteria, including two that were identified from the reference lists of initial selected articles (Figure 1). The six prediction models that were selected for further assessment were the Hepatic Steatosis Index (HSI) by Lee et al. [29], the Fatty Liver Disease Index (FLDI) by Fuyan et al. [30], the ZJU Index (ZJUI) by Wang et al. [31], the Framingham Steatosis Index (FSI) by Long et al. [32], the NAFLD Ridge Score by Yip et al. [33], and the NAFLD Scoring System by Lesmana et al. [34] (Table 1). The models were developed and published between 2010 and 2017. Two models

were from China, and there was one each from Korea, Indonesia, the United States (US), and Hong Kong. The risk algorithms of the selected models are presented in Table 2.

Validation cohort

A total of 404 individuals underwent liver elastography over the last three years. More than two-thirds of them (69.3%) were diagnosed with NAFLD. The most common unavailable or undocumented information included the HbA1c level (>15%), white blood cell count (>15%), BMI (10.9%), and AST level (1.7%). NAFLD was found to be more common among those who were older, female, of Malay ethnicity, and had a higher BMI and a larger waist circumference. They were also more likely to be diagnosed with diabetes mellitus (20.7%), hypertension (32.5%), and dyslipidaemia (34.6%). Their characteristics are summarized in Table 3.

Predictive performances

As demonstrated in the calibration plots (Figure 2), the predicted risks were closely clustered around the means in all the models. The FLDI, HSI, and ZJUI were found to have overestimated the NAFLD risk (intercept <0). All the models were overfitted (calibration slope <1) except for the NAFLD Ridge score. The FSI was found to be the best fit for the validation cohort (calibration slope = 0.84). Updating the models generally improved their calibration performance with a calibration line that was much closer to the ideal line (Figure 3). Nevertheless, the FLDI and NAFLD Ridge score did not show much improvement following the model update.

Figure 4 shows the AUROC for all the original prediction models. All models had a fair discrimination with an AUROC of above 0.7. The NAFLD Ridge Score yielded the lowest AUROC (0.717; 95% CI 0.662–0.772). Three models—the FSI, the FLDI, and the ZJUI—demonstrated the best discriminative ability with an AUROC of 0.783, 0.782, and 0.781, respectively. As updating the models did not change the ranking of the predicted risks of the patients, the AUROC of the models remained the same.

The sensitivity, specificity, likelihood ratios, positive and negative predictive values of the updated prediction models are shown in Table 4. By targeting the ten percent of the population that had values above the cut-offs, the HSI, FSI, and ZJUI were able to identify between 46% and 49% of the studied population who potentially had NAFLD. At 20% above the cut-off point, which is equal to the specificity of 80%, the sensitivity increased to 61%–67%. The FSI model was able to identify 47.5% and 66.8% individuals who had NAFLD by targeting the ten percent and 20% as the highest risk, respectively. Nevertheless, low specificity was seen in all prediction models when the sensitivity was above 90%. The NAFLD Ridge Score performed the worst with low sensitivity across all selected cut-offs.

Discussion

In order to identify a predictive model that could be beneficial for identifying NAFLD in Malaysia, this study evaluated the performance of six existing prediction models that apply routinely measured and

tested laboratory parameters in the country. The findings suggest that the FSI model by Long et al. [32], which was first developed based on the US population, has a better discrimination and calibration performance in predicting NAFLD in the Malaysian population.

The FSI model is simple and practical. It incorporates seven parameters, which are commonly tested and documented not only in hospitals but also in public healthcare centers across the country. This feature enables the early detection and management of NAFLD. The discovery that the three models with a high AUROC value (FSI, FLDI and ZJUI) are the ones that used similar predictors signifies a strong association between BMI, triglycerides, ALT/AST ratios, and an abnormal glucose level with the risk of developing NAFLD. Additionally, the FSI also considers the impact of age, sex, and hypertension on the development of NAFLD in the model's algorithm. Many previous studies found that NAFLD was more common in the elderly and males [35, 36] and that hypertension is an independent risk factor for NAFLD [37]. This could explain why the FSI model performs better when discriminating between individuals with and without NAFLD.

The first application of the original FSI model on the validation cohort demonstrated insufficient calibration. The predicted probabilities that were generated by this model were systematically too low with a calibration intercept of 2.05. Shen et al. reported a similar observation [38]. The systematic underestimation of the NAFLD risk in this study could partly be explained by the difference in incidence of NAFLD between the development cohort (317/1181, 26.8%) [32] and the validation cohort (280/404, 69.3%). To correct the miscalibration, a correction factor was added to the intercept of the original model. By combining the data that was included in the original model with a correction factor that was calculated from the current patient sample, the updated models were adjusted to the local population. Therefore, they yielded better calibration. Consequently, the calibration performance of the updated FSI model was improved with a new intercept that was closer to zero and a calibration line that was closer to the ideal line.

With a good predictive performance, as demonstrated by the updated FSI model in the current study, it is recommended that it is used as a simple screening tool in the local population to determine their risk of having NAFLD. If the model is used in the clinical practice, an empirical impact study on its value in improving the patient care, reducing the burden of the healthcare system, and enhancing the patient satisfaction is warranted.

The main strength of this study is that it is the first attempt to externally validate non-invasive prediction models for NAFLD and compare their performance using a Malaysian population. Incomplete information in the medical records of the validation cohort was minimal and was imputed to prevent biased results. The strength of this study also lies in the use of the CAP of the Fibroscan® for the diagnosis of NAFLD. The CAP has been shown to perform well and strongly agrees with the findings of liver biopsies on steatosis [39, 40]. A limitation of this study is that it only searched and examined prediction models that were found in articles from the PubMed database. Nevertheless, this database

was chosen because it contains a large number of articles that are written in English and covers all major journals across multiple clinical disciplines.

Conclusion

In conclusion, the updated FSI is favorable compared to other similar models to predict NAFLD in the Malaysian population. The application of routinely tested and documented parameters in this model makes it easy to use and practical.

Abbreviations

NAFLD – non-alcoholic fatty liver disease

ROC – receiver operating characteristic

AUROC – area under receiver operating characteristic

CAP – controlled attenuation parameter

FSI – Framingham Steatosis Index

HSI - Hepatic Steatosis Index

FLDI - Fatty Liver Disease Index

ALT – alanine aminotransferase

AST – aspartate aminotransferase

Declarations

Ethics approval and consent to participate

The conduct of this study was approved by the Medical Research and Ethics Committee of Ministry of Health Malaysia (NMRR-20–748-54587). The requirement for informed consent was waived because the data were retrospectively accessed.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

All authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contribution

MAMS created and designed the study, analyzed the data, and wrote the first draft of manuscript. HKC contributed to the manuscript organization and reviewed and edited the manuscript. SAS analyzed the data, checked the accuracy of the data analysis and contributed to the discussion. MRAH. oversee the conduct of the study, contributed to the discussion and reviewed the manuscript for intellectual content. All authors have full access to all data in the study and takes responsibility for the integrity of the data. All authors were also involved in the writing of the manuscript and approved the manuscript's final version.

Acknowledgements

The authors would like to thank Mohd Ammar Dzakirin Md Mansor, Siti Maisarah Md Ali and Mohamad Faiz Mustafa for their kind help during data collection. We also thank the Director General of Health Malaysia for his permission to publish this article.

References

1. Goh GBB, McCullough AJ. Natural history of nonalcoholic fatty liver disease. *Dig Dis Sci.* 2016;61(5):1226–1233.
2. 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 Baltim Md.* 2016;64(1):73–84.
3. Kang Y, Park S, Kim S, Koh H. Estimated prevalence of adolescents with nonalcoholic fatty liver disease in Korea. *J Korean Med Sci.* 2018;33(14): e109.
4. Wu Y, Zheng Q, Zou B, et al. The epidemiology of NAFLD in mainland China with analysis by adjusted gross regional domestic product: a meta-analysis. *Hepatology Int.* 2020;14(2):259–269.
5. Goh GBB, Kwan C, Lim SY, et al. Perceptions of non-alcoholic fatty liver disease-an Asian community-based study. *Gastroenterol Rep.* 2016;4(2):131–135.
6. Kumar R, Priyadarshi RN, Anand U. Non-alcoholic fatty liver disease: growing burden, adverse outcomes, and associations. *J Clin Transl Hepatol.* 2020;8(1):76–86.

7. Kanwal F, Kramer JR, Mapakshi S, et al. Risk of hepatocellular cancer in patients with non-alcoholic fatty liver disease. *Gastroenterology*. 2018;155(6):1828–1837.
8. Kishanifarahani Z, Ahadi M, Kazeminejad B, et al. Inter-observer variability in histomorphological evaluation of non-neoplastic liver biopsy tissue and impact of clinical information on final diagnosis in Shahid Beheshti University of Medical Sciences Affiliated Hospitals. *Iran J Pathol*. 2019;14(3):243–247.
9. European Association for the Study of the Liver. EASL-ALEH Clinical Practice Guidelines: non-invasive tests for evaluation of liver disease severity and prognosis. *J Hepatol*. 2015;63(1):237–264.
10. Mueller M, Kratzer W, Oeztuerk S, et al. Percutaneous ultrasonographically guided liver punctures: an analysis of 1961 patients over a period of ten years. *BMC Gastroenterol*. 2012;12:173.
11. Khammas ASA, Hassan HA, Salih SQM, et al. Prevalence and risk factors of sonographically detected non-alcoholic fatty liver disease in a screening centre in Klang Valley, Malaysia: an observational cross-sectional study. *Porto Biomed J*. 2019;4(2):e31.
12. Cheah WL, Lee PY, Chang CT, Mohamed HJ, Wong SL. Prevalence of ultrasound diagnosed nonalcoholic fatty liver disease among rural indigenous community of Sarawak and its association with biochemical and anthropometric measures. *Southeast Asian J Trop Med Public Health*. 2013;44(2):309–317.
13. Hassan HA, Koh CT, Apandi LM, Suppiah S, Rahim EA. Prevalence of fatty liver changes on non-contrast enhanced computed tomography and its associated risk factors. *Int J Public Health Clin Sci*. 2019;6(4):68–78.
14. Chan WK, Tan ATB, Vethakkan SR, Tah PC, Vijayanathan A, Goh KL. Non-alcoholic fatty liver disease in diabetics-prevalence and predictive factors in a multiracial hospital clinic population in Malaysia. *J Gastroenterol Hepatol*. 2013;28(8):1375–1383.
15. Magosso E, Ansari MA, Gopalan Y, et al. Prevalence of non-alcoholic fatty liver in a hypercholesterolemic population of northwestern peninsular Malaysia. *Southeast Asian J Trop Med Public Health*. 2010;41(4):936–942.
16. Suppiah S, Lee RMC, Sazali NS, Hassan HA. Non-alcoholic fatty liver disease in metabolic syndrome patients in Serdang Hospital: quantification by contrast-enhanced computed tomography. *Malays J Med Health Sci*. 2016;12(1):9–18.
17. Wong VWS, Chan WK, Chitturi S, et al. Asia-Pacific Working Party on non-alcoholic fatty liver disease guidelines 2017-Part 1: definition, risk factors and assessment. *J Gastroenterol Hepatol*. 2018;33(1):70–85.
18. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.
19. Karlas T, Petroff D, Sasso M, et al. Individual patient data meta-analysis of controlled attenuation parameter (CAP) technology for assessing steatosis. *J Hepatol*. 2017;66(5):1022–1030.
20. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med*. 2016;35(2):214–226.

21. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol*. 2005;58(5):475–483.
22. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Med Res Methodol*. 2014;14(1):75.
23. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925-1931.
24. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiol Camb Mass*. 2010;21(1):128–138.
25. Janssen KJM, Vergouwe Y, Kalkman CJ, Grobbee DE, Moons KGM. A simple method to adjust clinical prediction models to local circumstances. *Can J Anaesth*. 2009;56(3):194–201.
26. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models II: external validation, model updating, and impact assessment. *Heart Br Card Soc*. 2012;98(9):691–698.
27. Swets JA. Measuring the accuracy of diagnostic systems. *Science*. 1988;240(4857):1285–1293.
28. Team RC. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2014.
29. Lee JH, Kim D, Kim HJ, et al. Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. *Dig Liver Dis*. 2010;42(7):503–508.
30. Fuyan S, Jing L, Wenjun C, et al. Fatty liver disease index: a simple screening tool to facilitate diagnosis of nonalcoholic fatty liver disease in the Chinese population. *Dig Dis Sci*. 2013;58(11):3326–3334.
31. Wang J, Xu C, Xun Y, et al. ZJU index: a novel model for predicting nonalcoholic fatty liver disease in a Chinese population. *Sci Rep*. 2015;5:16494.
32. Long MT, Pedley A, Colantonio LD, et al. Development and validation of the Framingham Steatosis Index to identify persons with hepatic steatosis. *Clin Gastroenterol Hepatol*. 2016;14(8):1172–1180.
33. Yip TCF, Ma AJ, Wong VWS, et al. Laboratory parameter-based machine learning model for excluding non-alcoholic fatty liver disease (NAFLD) in the general population. *Aliment Pharmacol Ther*. 2017;46(4):447–456.
34. Lesmana CRA, Pakasi LS, Inggriani S, Aidawati ML, Lesmana LA. Development of non-alcoholic fatty liver disease scoring system among adult medical check-up patients: a large cross-sectional and prospective validation study. *Diabetes Metab Syndr Obes Targets Ther*. 2015;8:213–218.
35. Gan L, Chitturi S, Farrell GC. Mechanisms and implications of age-related changes in the liver: nonalcoholic fatty liver disease in the elderly. *Curr Gerontol Geriatr Res*. 2011;2011:831536.
36. Golabi P, Paik J, Reddy R, Bugianesi E, Trimble G, Younossi ZM. Prevalence and long-term outcomes of non-alcoholic fatty liver disease among elderly individuals from the United States. *BMC Gastroenterol*. 2019;19(1):56.

37. Hu XY, Li Y, Li LQ, et al. Risk factors and biomarkers of non-alcoholic fatty liver disease: an observational cross-sectional population survey. *BMJ Open*. 2018;8(4):e019974.
38. Shen YN, Yu MX, Gao Q, et al. External validation of non-invasive prediction models for identifying ultrasonography-diagnosed fatty liver disease in a Chinese population. *Medicine (Baltimore)*. 2017;96(30):e7610.
39. da Silva L de CM, de Oliveira JT, Tochetto S, de Oliveira CPMS, Sigrist R, Chammas MC. Ultrasound elastography in patients with fatty liver disease. *Radiol Bras*. 2020;53(1):47–55.
40. Jun BG, Park WY, Park EJ, et al. A prospective comparative assessment of the accuracy of the FibroScan in evaluating liver steatosis. *PloS One*. 2017;12(8):e0182784.

Tables

Table 1 Characteristics of included prediction models for external validation

Model name	Study period	Study design	No. cases / total	Diagnosis NAFLD	Predictors
Hepatic Steatosis Index	2006	Cross-sectional case-control	5769 / 17,539	Ultrasonography	BMI, ALT/AST ratio, diabetes.
Fatty Liver Disease Index	2012	Cross-sectional case-control	3671 / 16,044	Ultrasonography	BMI, triglycerides, ALT/AST ratio, hyperglycaemia.
ZJU Index	2014	Cross-sectional case-control	4801 / 13,729	Ultrasonography	BMI, triglycerides, ALT/AST ratio, fasting plasma glucose.
NAFLD Score System	2013	Cross-sectional study	538 / 1054	Ultrasonography	Age, sex, BMI, triglycerides, HDL cholesterol, ALT.
Framingham Steatosis Index	2008–2011	Cross-sectional study	317 / 1181	Computed tomography	Age, sex, BMI, triglycerides, hypertension, diabetes, ALT/AST ratio.
NAFLD Ridge Score	2008–2010	Cross-sectional study	264 / 922	Proton-magnetic resonance spectroscopy	ALT, HDL cholesterol, triglycerides, HbA1c, WBC count, hypertension.
NAFLD, non-alcoholic fatty liver disease; ALT, alanine aminotransferase; AST, aspartate aminotransferase; WBC, white blood cell; BMI, body mass index; HDL, high density lipoprotein.					

Table 2 Included model algorithms for predicting non-alcoholic fatty liver disease

Model name, first author (year)	The probability of developing Non-Alcoholic Fatty Liver Disease was calculated as $e^{lp} / (1 + e^{lp})$, where:
Hepatic Steatosis Index, Lee (2010)	$lp = -9.960 + 0.315 (\text{BMI}) + 2.421 (\text{ALT/AST ratio}) + 0.630 (\text{diabetes [presence=1, absence=0]})$
Fatty Liver Disease Index, Fuyan (2013)	$lp = 0.3413 (\text{BMI}) + 0.3134 (\text{triglycerides, mmol/L}) + 0.9499 (\text{ALT/AST ratio}) + 0.6710 (\text{hyperglycaemia [presence=1, absence=0]})$
ZJU Index, Wang (2015)	$lp = -10.52 + 0.317 (\text{BMI}) + 0.875 (\text{ALT/AST ratio}) + 0.227 (\text{fasting plasma glucose, mmol/L}) + 0.362 (\text{triglycerides, mmol/L})$
NAFLD Score System, Lesmana (2015)	$lp = -2.568 + 0.478 [1.9 (\text{sex [female = 0, male = 1]}) + 2.1(\text{age [>35 = 1, ≤35 = 0]}) + 3.6 (\text{BMI [≥25 = 1, <25 = 0]}) + 1.6 (\text{triglycerides, mg/dl [≥150 = 1, <150 = 0]}) + 1.0 (\text{HDL, mg/dL [<40 \& male or <50 \& female] = 1}) + 2.0 (\text{ALT, U/L [≥35 = 1, <35 = 0]})]$
Framingham Steatosis Index, Long (2016)	$lp = -7.981 + 0.011 (\text{age}) - 0.146 (\text{sex [female = 1, male = 0]}) + 0.173 (\text{BMI}) + 0.007 (\text{triglycerides, mg/dl}) + 0.593 (\text{hypertension [yes = 1, no = 0]}) + 0.789 (\text{diabetes [yes = 1, no = 0]}) + 1.1 (\text{ALT/AST ratio} \geq 1.33 [\text{yes = 1, no = 0}])$
NAFLD Ridge Score, Yip (2017)	$lp = -0.614 + 0.007 (\text{ALT, IU/L}) - 0.214 (\text{HDL, mmol/L}) + 0.053 (\text{triglyceride, mmol/L}) + 0.144 (\text{HbA1c, \%}) + 0.032 (\text{WBC count, } \times 10^9/\text{L}) + 0.132 (\text{hypertension [yes = 1, no = 0]})$
lp, linear prediction; ALT, alanine aminotransferase; AST, aspartate aminotransferase; WBC, white blood cell; BMI, body mass index; HDL, high density lipoprotein.	

Table 3 Baseline characteristics of the validation cohort

Variables	Missing values, n (%)	Validation cohort †				
		Overall (n = 404)	NAFLD (n = 280)		Non-NAFLD (n = 124)	
Age (years), mean (SD)	0 (0.0)	50.0 (13.27)	51.2	(12.70)	47.1	(14.09)
Ethnicity, n (%)	0 (0.0)					
Malay		267 (66.1)	174	(62.1)	93	(75.0)
Non-Malay		137 (33.9)	106	(37.9)	31	(25.0)
Gender, n (%)	0 (0.0)					
Male		169 (41.8)	127	(45.4)	42	(33.9)
Female		235 (58.2)	153	(54.6)	82	(66.1)
BMI (kg/m ²), mean (SD)	44 (10.9)	29.1 (6.09)	30.0	(6.29)	25.9	(4.01)
Diabetes mellitus, n (%)	0 (0.0)	75 (18.6)	58	(20.7)	17	(13.7)
Hypertension, n (%)	0 (0.0)	115 (28.5)	91	(32.5)	24	(19.4)
Dyslipidaemia, n (%)	0 (0.0)	136 (33.7)	97	(34.6)	39	(31.4)
ALT (U/L), mean (SD)	0 (0.0)	37.2 (32.0)	41.7	(30.20)	27.1	(33.58)
AST (U/L), mean (SD)	7 (1.7)	30.1 (27.25)	30.1	(14.24)	30.2	(45.25)
FBS (mmol/L), mean (SD)	0 (0.0)	5.8 (2.40)	6.0	(2.66)	5.3	(1.58)
TG (mmol/L), mean (SD)	0 (0.0)	1.7 (1.16)	1.8	(1.29)	1.3	(0.63)
HDL (mg/dL), mean (SD)	0 (0.0)	1.4 (0.36)	1.3	(0.34)	1.5	(0.38)
HbA1C (%), mean (SD)	71 (17.6)	6.3 (1.60)	6.4	(1.65)	6	(1.36)
WBC count (x10 ⁹ /L), mean (SD)	67 (16.6)	7.5 (2.43)	7.5	(2.45)	7.4	(2.39)

† Original data (not imputed).

NAFLD, non-alcoholic fatty liver disease; SD, standard deviation; BMI, body mass index; ALT, alanine aminotransferase; AST, aspartate aminotransferase; FBS, fasting blood sugar; TG, triglycerides; HDL, high density lipoprotein; WBC, white blood cell.

Table 4 Discriminatory performance measures of updated prediction models for the risk of developing NAFLD

	Hepatic Steatosis Index	Fatty Liver Disease Index	ZJU Index	Framingham Steatosis Index	NAFLD Ridge Score	NAFLD Scoring System
10% above cut-off point						
Sensitivity	46.8	43.6	48.9	47.5	31.1	45.7
Specificity	90.0	90.0	90.0	90.0	90.0	90.0
LR +	4.68	4.36	4.89	4.75	3.11	4.57
LR -	0.59	0.63	0.57	0.58	0.77	0.60
PPV	91.0	90.4	91.9	91.1	87.9	90.8
NPV	42.7	41.3	43.9	43	36.7	42.2
20% above cut-off point						
Sensitivity	62.5	60	60.7	66.8	46.4	58.2
Specificity	80.0	80.0	80.0	80.0	80.0	80.0
LR +	3.13	3.00	3.04	3.34	2.32	2.91
LR -	0.47	0.50	0.49	0.42	0.67	0.52
PPV	87.5	87	87.2	88.2	83.9	86.2
NPV	48.5	46.9	47.4	51.6	39.8	45.6
80% above cut-off point						
Sensitivity	93.6	96.4	96.1	96.1	91.8	92.9
Specificity	20.0	20.0	20.0	20.0	20.0	20.0
LR +	1.17	1.21	1.20	1.20	1.15	1.16
LR -	0.32	0.18	0.20	0.20	0.41	0.36
PPV	72.6	73.2	73.1	73.1	72.2	72.2
NPV	58.1	71.4	69.4	69.4	52.1	54.5
90% above cut-off point						
Sensitivity	97.5	98.2	98.2	96.4	96.1	98.2
Specificity	10.5	10.5	10.5	10.5	10.5	10.0
LR +	1.09	1.10	1.10	1.08	1.07	1.09

LR -	0.24	0.17	0.17	0.34	0.37	0.18
PPV	71.1	71.2	71.2	70.9	70.8	71.1
NPV	65.0	72.2	72.2	56.5	54.2	70.6

NAFLD, non-alcoholic fatty liver disease; LR+, positive likelihood ratio; LR-, negative likelihood ratio; PPV, positive predictive value; NPV, negative predictive value.

Figures

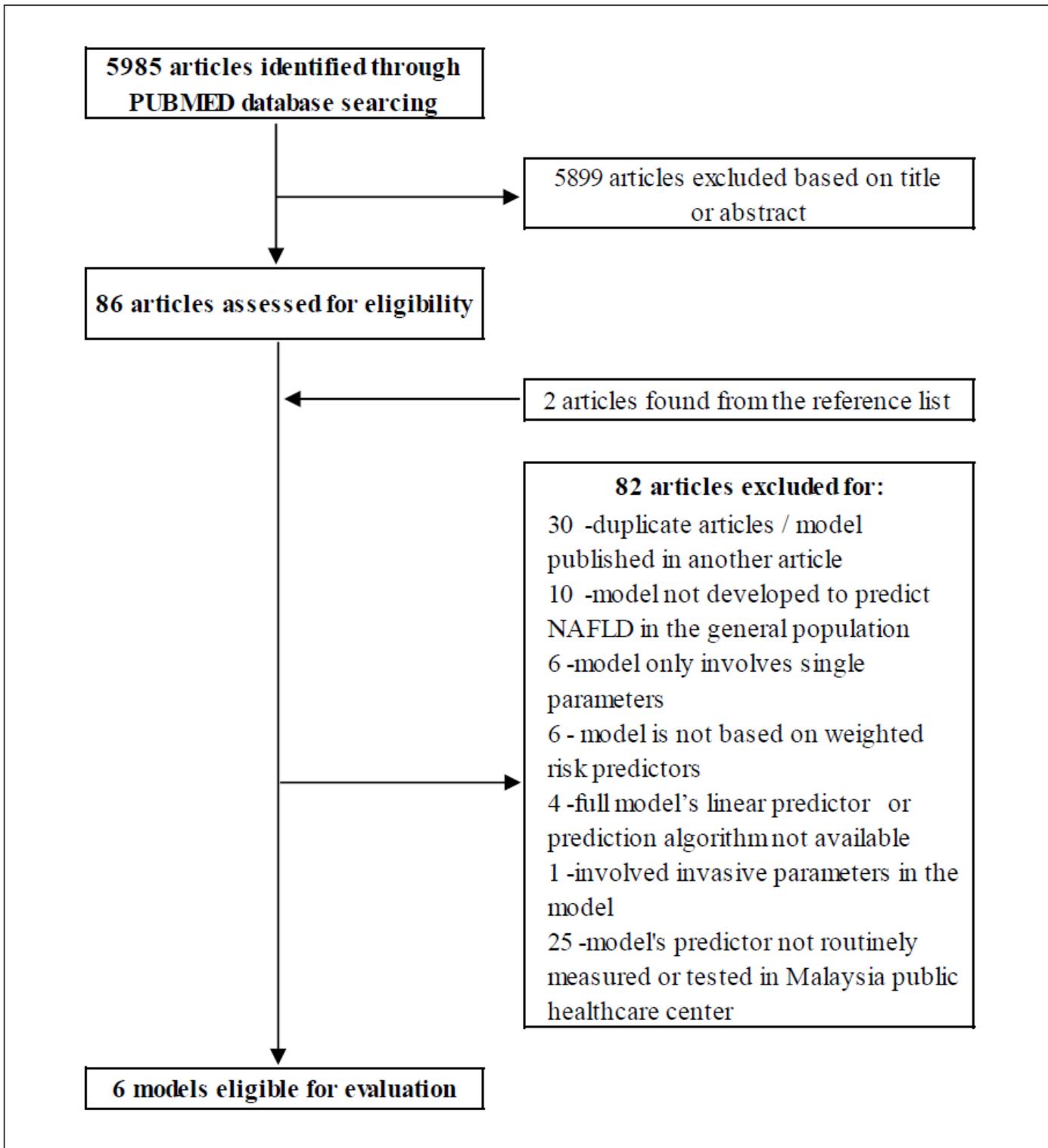


Figure 1

Flowchart of model selection.

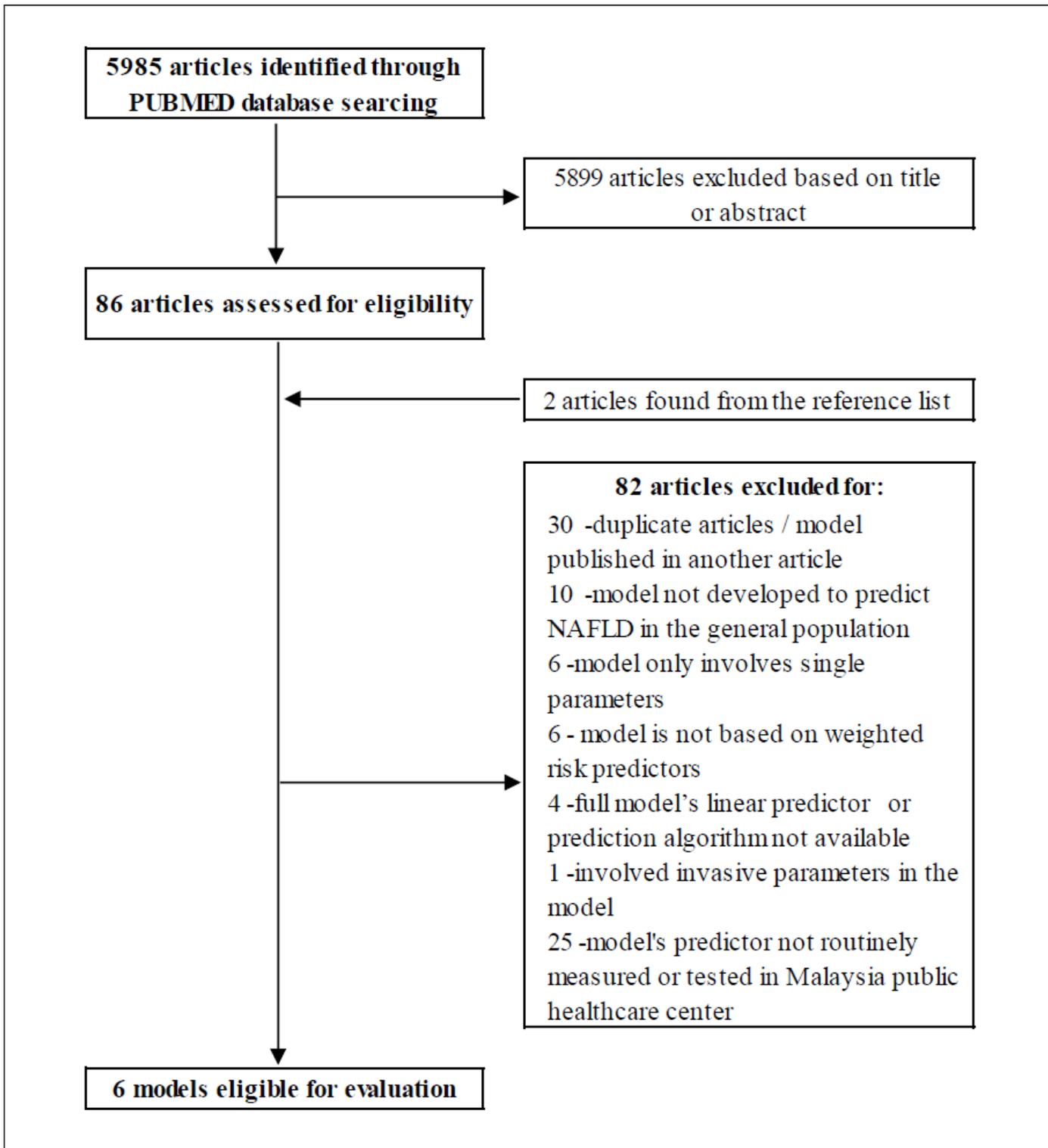


Figure 1

Flowchart of model selection.

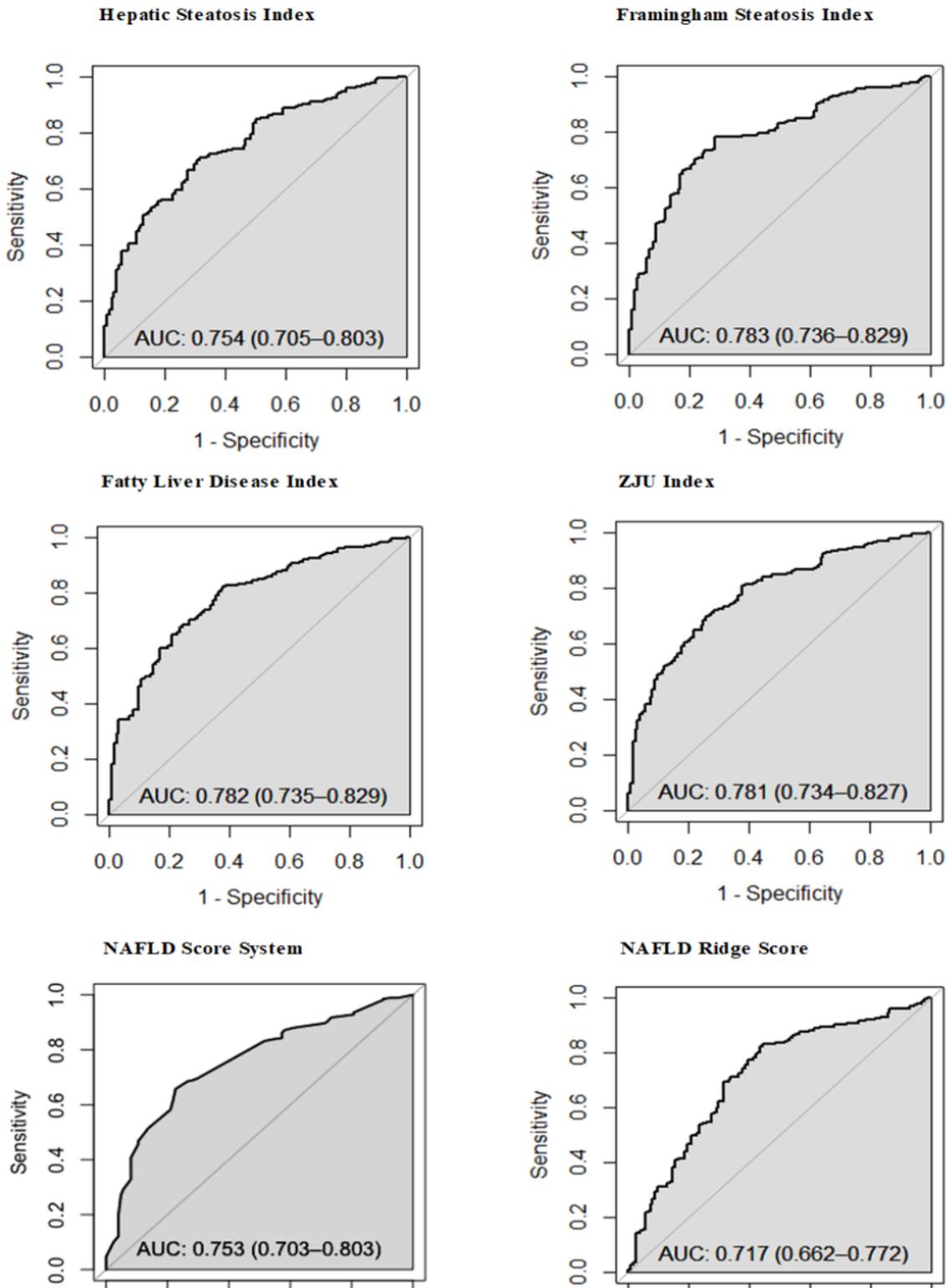


Figure 2

Calibration plot of the existing prediction models for NAFLD. The grey dashed line is the reference line with intercept = zero and slope = one (perfect calibration). The triangles correspond to grouped predicted risks with 95% confidence intervals (vertical lines).

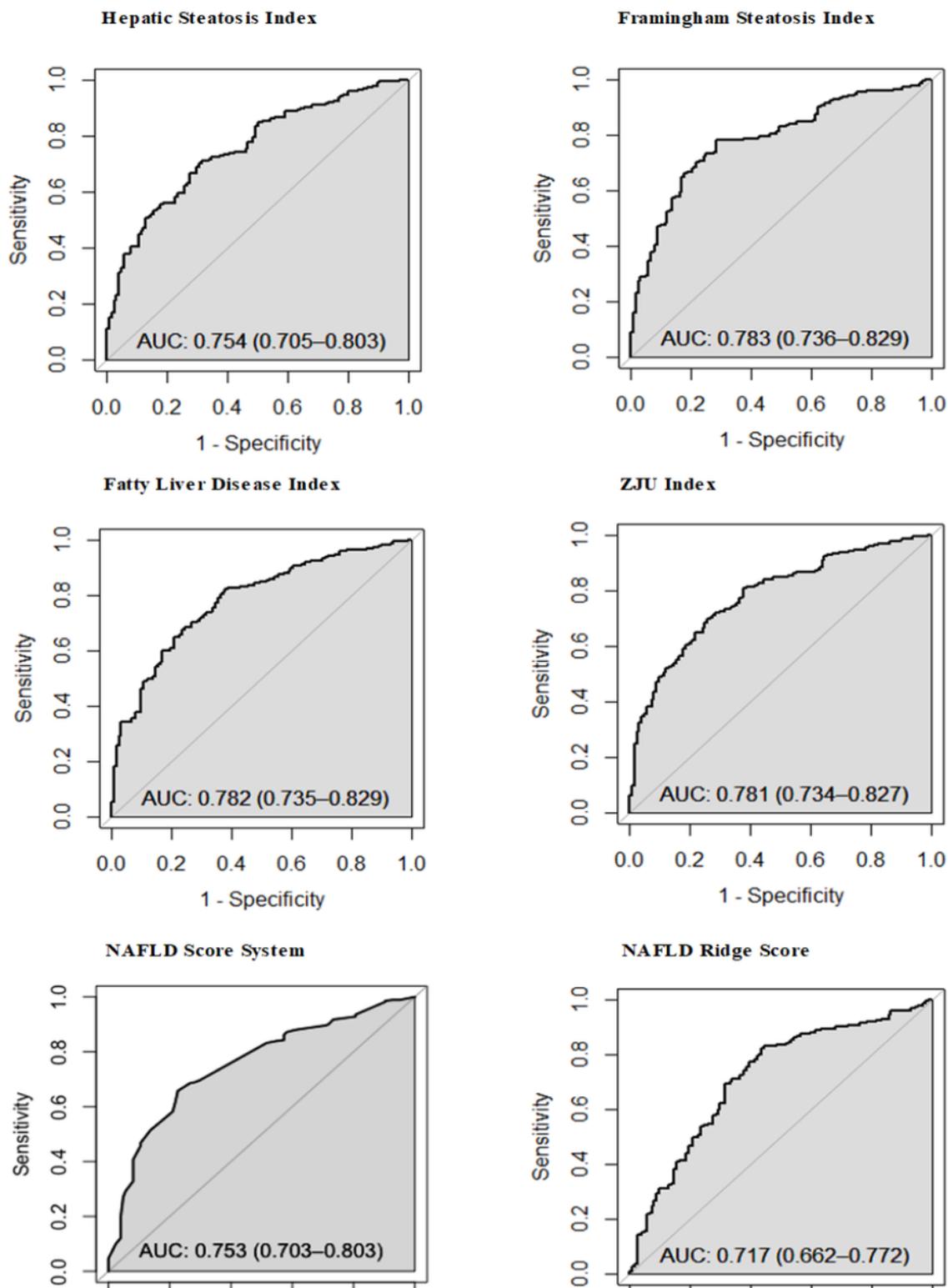


Figure 2

Calibration plot of the existing prediction models for NAFLD. The grey dashed line is the reference line with intercept = zero and slope = one (perfect calibration). The triangles correspond to grouped predicted risks with 95% confidence intervals (vertical lines).

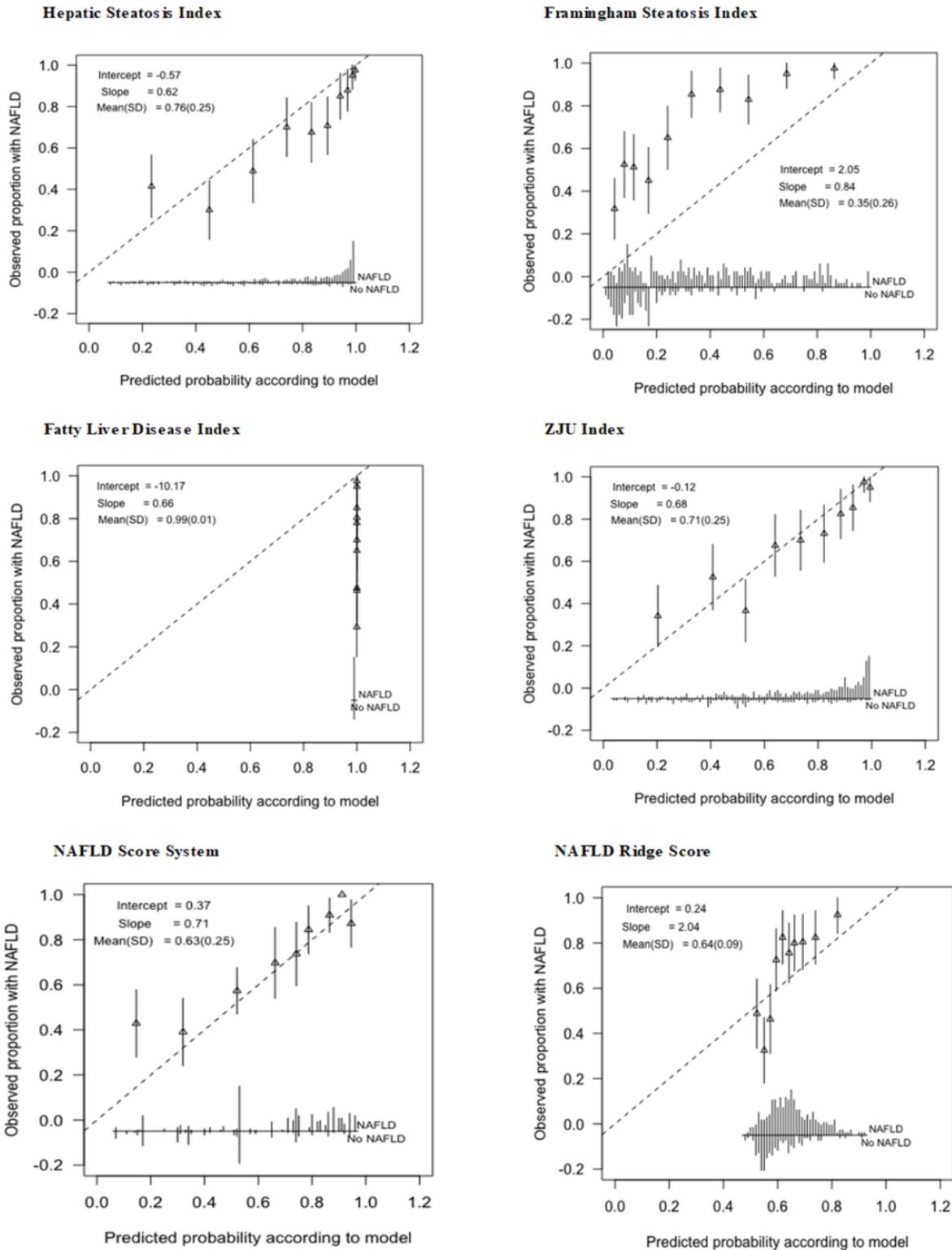


Figure 3

Calibration plot of the updated models for NAFLD. The grey dashed line is the reference line with intercept = zero and slope = one (perfect calibration). The triangles correspond to grouped predicted risks with 95% confidence intervals (vertical lines). CF = correction factor.

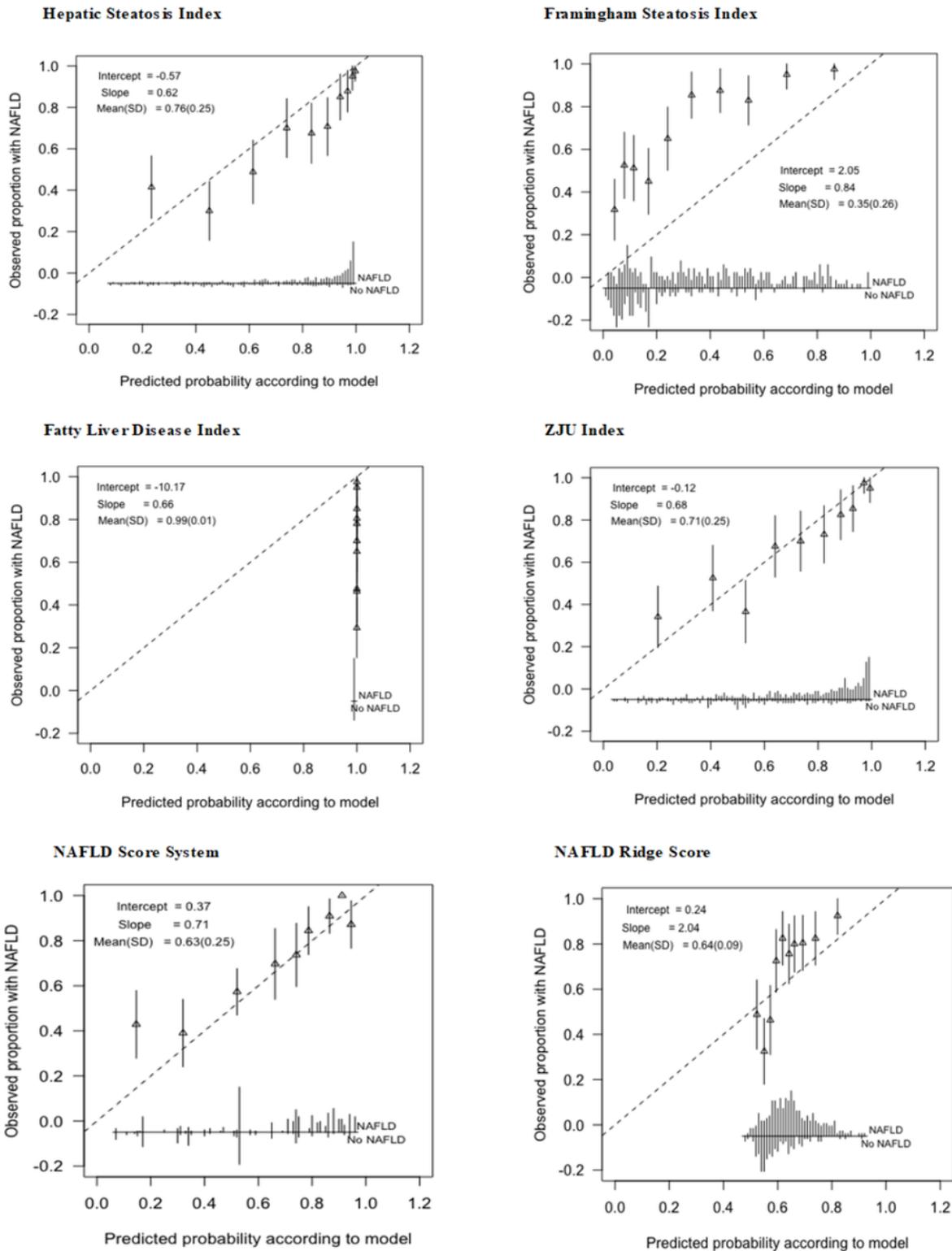


Figure 3

Calibration plot of the updated models for NAFLD. The grey dashed line is the reference line with intercept = zero and slope = one (perfect calibration). The triangles correspond to grouped predicted risks with 95% confidence intervals (vertical lines). CF = correction factor.

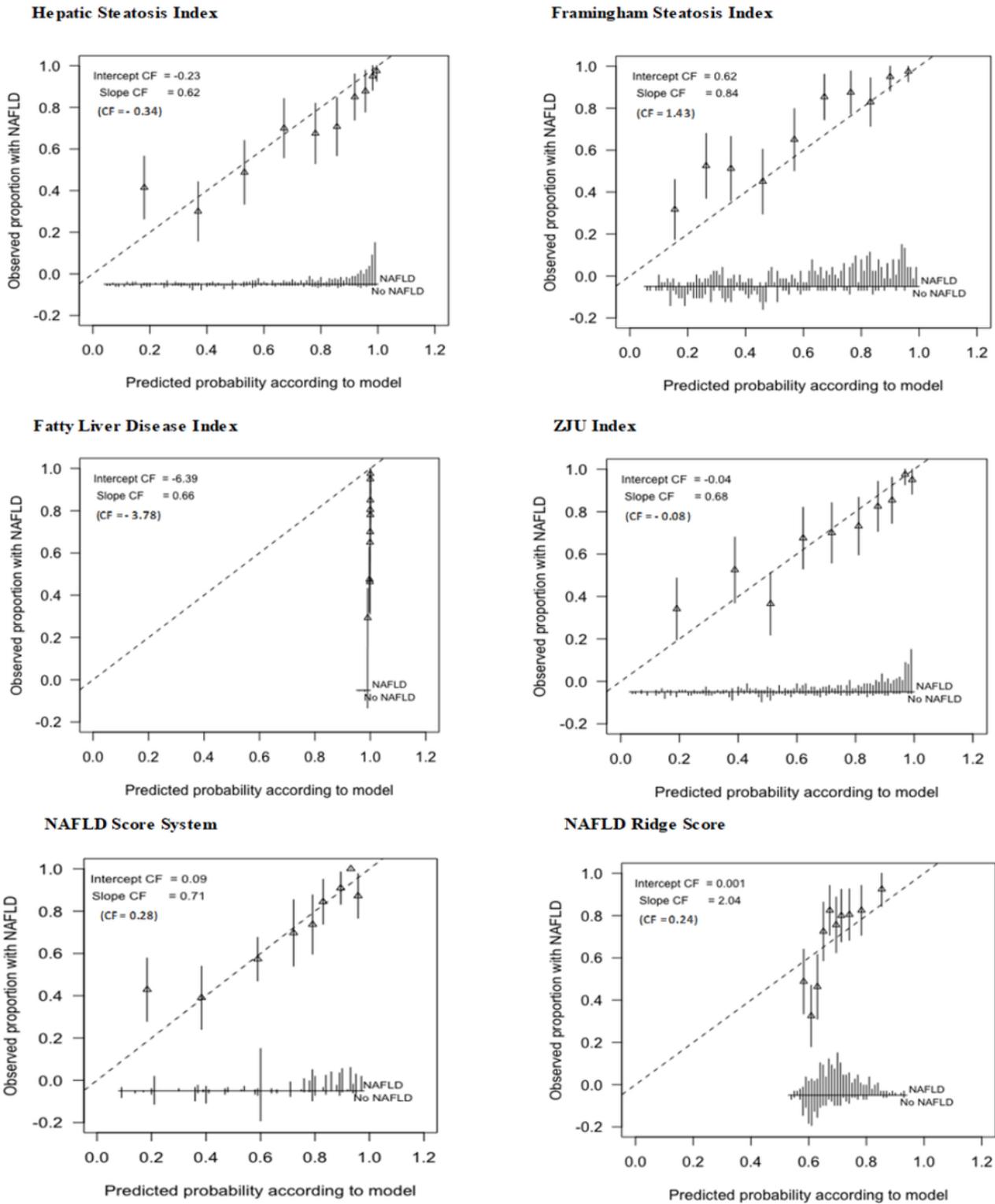


Figure 4

Area under the receiver operating characteristic curve of prediction models.

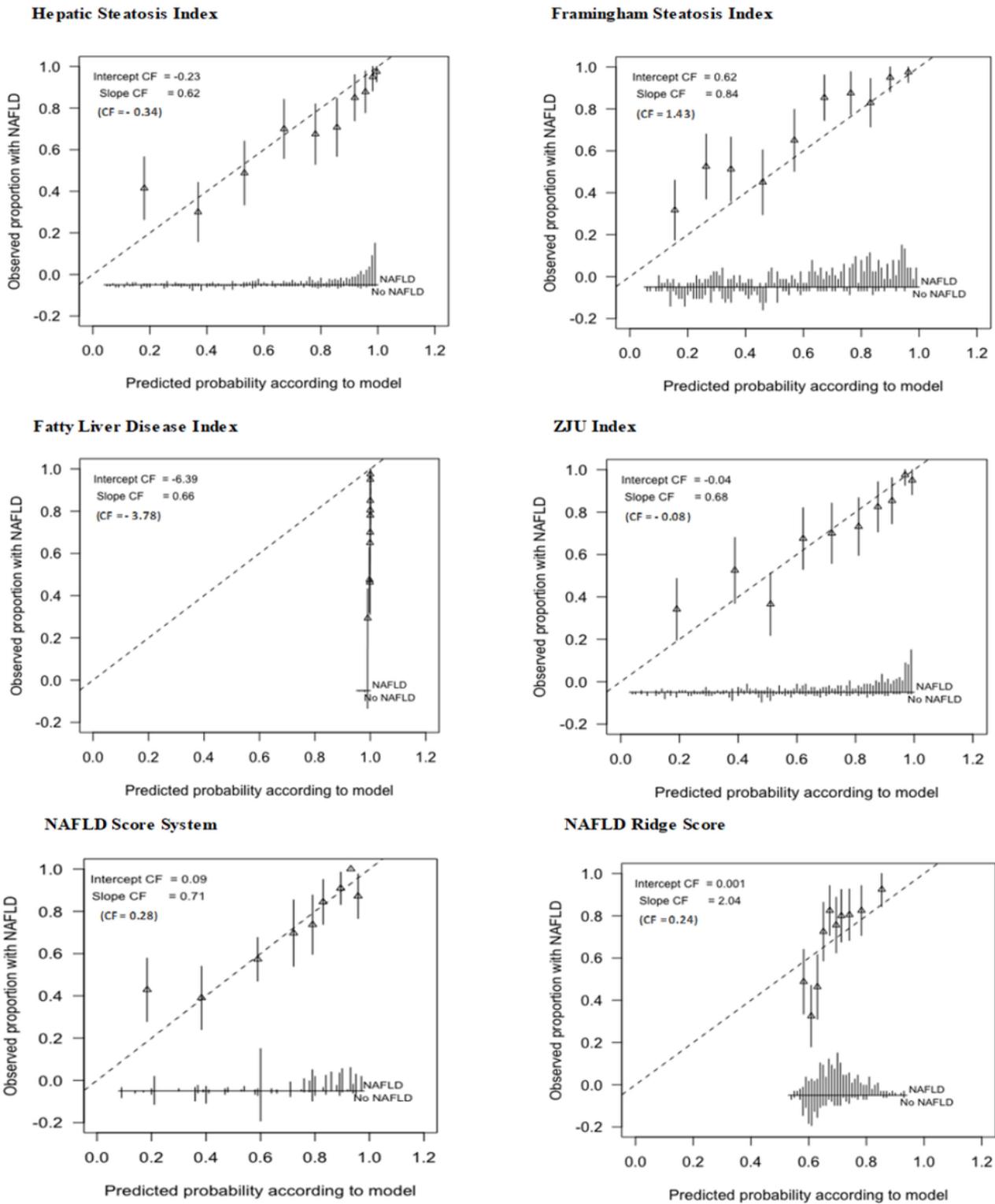


Figure 4

Area under the receiver operating characteristic curve of prediction models.