Cureus
Part of SPRINGER NATURE

Open Access Original Article

Development and Validation of a Nomogram for Identifying Low Appendicular Muscle Mass in Patients With Nonalcoholic Fatty Liver Disease: A Cross-Sectional Study

Review began 10/12/2026
Review ended 10/27/2026
Published 11/01/2026

© Copyright 2026
Pang et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.118926

Yujiang Pang [1], Wenluo Zhang [1], Mengni Jing [1], Yifan Ling [1], Zuorui Zhu [1], Jie Zhou [2], Zhenhua Hu [3]

1. Department of Hepatobiliary and Pancreatic Surgery, Department of Surgery, The Fourth Affiliated Hospital of School of Medicine, Zhejiang University International School of Medicine, Yiwu, CHN 2. Department of Hepatobiliary and Pancreatic Surgery, Department of Surgery, The First Affiliated Hospital School of Medicine, Zhejiang University International School of Medicine, Hangzhou, CHN 3. Department of Surgical Oncology, The First Affiliated Hospital School of Medicine, Zhejiang University International School of Medicine, Hangzhou, CHN

Corresponding author: Zhenhua Hu, huzhenh@zju.edu.cn

Abstract

Background: Low appendicular muscle mass is a component of sarcopenia that can be measured using dual-energy X-ray absorptiometry. This study developed and internally validated a nomogram for identifying prevalent low appendicular muscle mass in adults with nonalcoholic fatty liver disease (NAFLD).

Methods: We analyzed 1,571 participants aged 18-59 years from the National Health and Nutrition Examination Survey (NHANES) 2011-2018, randomly divided into training (n = 1,099) and testing (n = 472) sets. Low appendicular muscle mass was defined using sex-specific thresholds for appendicular skeletal muscle mass divided by body mass index (BMI). Least absolute shrinkage and selection operator (LASSO) and logistic regression were used for model development. Discrimination, calibration, decision curves, 1,000 bootstrap resamples, and comparisons with simpler models were evaluated. Prediction modeling and performance assessment were unweighted.

Results: The model included age, race/ethnicity, BMI, height, triglycerides, phosphorus, and chloride. The area under the receiver operating characteristic curve (AUC) was 0.838 (95% CI: 0.808-0.868) in training and 0.812 (95% CI: 0.757-0.868) in testing. The optimism-corrected training AUC was 0.826. At a training-derived probability cut-off of 0.1838, testing sensitivity and specificity were 65.28% and 78.25%. The testing calibration intercept and slope were 0.0010 and 0.9725. Adding laboratory variables to age, race/ethnicity, BMI, and height did not improve testing discrimination (AUC: 0.812 versus 0.813; DeLong p = 0.947). Decision-curve differences varied by threshold probability.

Conclusions: The nomogram showed internal discrimination for prevalent low appendicular muscle mass in selected adults with NAFLD. Interpretation is limited by the structural relationship between anthropometric predictors and the BMI-adjusted outcome. Incremental benefit from the laboratory variables was not established in the testing set. External and prospective clinical validation are required before implementation.

Categories: Nutrition, Gastroenterology
Keywords: identification, low appendicular muscle mass, nhanes, nomogram, nonalcoholic fatty liver disease (nafld)

Introduction

Nonalcoholic fatty liver disease (NAFLD) is common worldwide [1] and is associated with liver fibrosis and extrahepatic morbidity [2,3]. Skeletal muscle dysfunction is relevant to its systemic manifestations [4]. Current guidelines use the term metabolic dysfunction-associated steatotic liver disease (MASLD), defined by hepatic steatosis and at least one cardiometabolic risk factor [5]. We retain NAFLD terminology because the original study eligibility criteria were based on NAFLD definitions; participants were not reclassified as having MASLD.

Sarcopenia encompasses abnormalities in muscle mass and strength, and contemporary definitions emphasize muscle function and quantity [6-8]. Low appendicular muscle mass is one measurable component and should be distinguished from a complete clinical diagnosis. Studies have associated sarcopenia with advanced fibrosis in NAFLD [9]. Impaired muscle glucose uptake, insulin resistance, inflammatory signaling, and altered liver-muscle interactions provide a biological context for this association [10-12]. Observational studies have also linked sarcopenia with NAFLD, steatohepatitis, fibrosis, and mortality [13-15]. These findings support investigation of muscle assessment in NAFLD while leaving the clinical value of individual identification models to be established.

How to cite this article
Pang Y, Zhang W, Jing M, et al. (November 01, 2026) Development and Validation of a Nomogram for Identifying Low Appendicular Muscle Mass in Patients With Nonalcoholic Fatty Liver Disease: A Cross-Sectional Study. Cureus 18(11): e118926. DOI 10.7759/cureus.118926

This study aimed to develop and internally validate a nomogram for identifying prevalent low appendicular muscle mass in adults with NAFLD. Its proposed setting is hepatology or gastroenterology clinics, where clinicians could use existing anthropometric and laboratory results to identify patients who may warrant further assessment of muscle mass, strength, and physical performance. This is a proposed use requiring prospective evaluation; the study did not test model-guided referral or patient outcomes.

Materials And Methods

Study population

This cross-sectional study used the public-use National Health and Nutrition Examination Survey (NHANES) 2011-2018 data, collected by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC) [16]. The initial survey sample comprised 39,156 participants. NAFLD was identified using the Fatty Liver Index (FLI) ⩾60 in 2011-2016 [17] and controlled attenuation parameter (CAP) ⩾248 dB/m in 2017-2018 [18]. For the latter cycle, participants with missing CAP or liver stiffness measurements, incomplete elastography examinations, or liver stiffness ⩾8.2 kPa were excluded. The liver stiffness threshold was intended to limit inclusion of potentially significant fibrosis [19] and was applied only in 2017-2018.

The FLI was calculated as

$$100 \times exp(L)/[1 + exp(L)]$$

where L = 0.953 × ln(triglycerides) + 0.139 × BMI + 0.718 × ln(GGT) + 0.053 × waist circumference − 15.745 [17]. The formula used triglycerides in mg/dL, gamma-glutamyl transferase (GGT) in U/L, BMI in kg/m^2, and waist circumference in cm. Triglycerides are reported in mmol/L elsewhere in this manuscript, using mmol/L = mg/dL × 0.01129.

Participants with hepatitis B, hepatitis C, autoimmune hepatitis, or liver cancer were excluded using NHANES serological measurements and medical-history questionnaires. Excessive alcohol intake, assessed from dietary recall, was defined as >20 g/day in men and >10 g/day in women and was also excluded. Figure *1* summarizes selection: 4,907 participants met the NAFLD eligibility criteria, and 2,422 had DXA data. Excluding 701 participants with missing smoking, recreational activity, diabetes, or blood pressure data left 1,721 participants; a further 150 with missing covariate data were excluded, yielding 1,571 complete cases. The final sample and both split datasets comprised adults aged 18-59 years, with no participants younger than 18 years. Original NHANES protocols received NCHS Ethics Review Board approval, and written informed consent was obtained [20]. The completed TRIPOD+AI reporting checklist is provided in the Appendices (Supplementary Table S6).

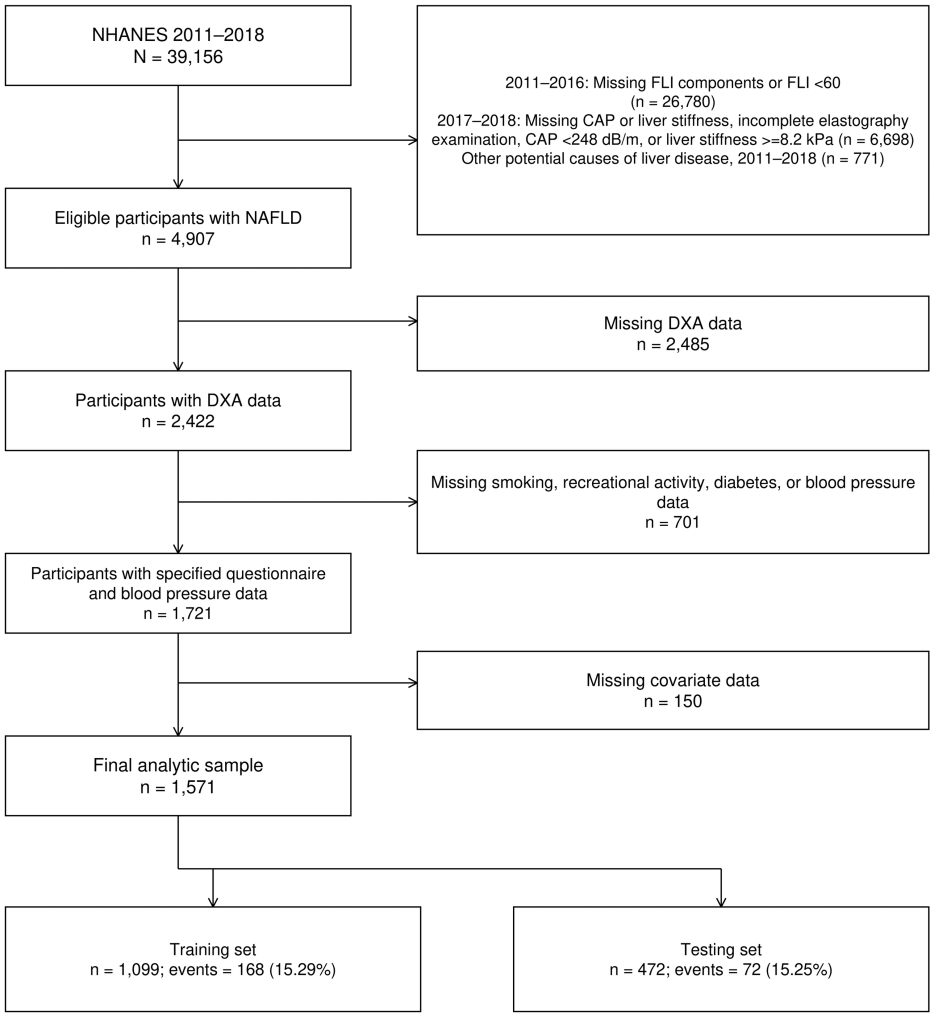


FIGURE 1: Flowchart of participant selection from NHANES 2011-2018.

Abbreviations: NHANES, National Health and Nutrition Examination Survey; NAFLD, nonalcoholic fatty liver disease; DXA, dual-energy X-ray absorptiometry; FLI, Fatty Liver Index; CAP, controlled attenuation parameter; BMI, body mass index

Measurements and definition of low appendicular muscle mass

Appendicular skeletal muscle mass (ASM) was estimated as the sum of DXA-derived lean mass in both arms and legs. The BMI-adjusted ASM index (ASMI) was calculated as ASM in kilograms divided by BMI in kg/m². Low appendicular muscle mass was defined as ASMI <0.789 in men and <0.512 in women, using the Foundation for the National Institutes of Health (FNIH) Sarcopenia Project cutoffs [21]. This outcome measures a muscle-mass phenotype; muscle strength and physical performance were not included in its definition [6,7].

Variable estimation

Candidate predictors were obtained from NHANES demographic, examination, questionnaire, and laboratory files and included age, sex, race/ethnicity, body size, blood pressure, physical activity, smoking, complete blood count, and biochemistry. Height, weight, and waist circumference were measured using standardized procedures, and BMI was calculated as weight in kilograms divided by height in meters squared. Self-reported NHANES race and Hispanic origin categories were Mexican American, other Hispanic, non-Hispanic White, non-Hispanic Black, non-Hispanic Asian, and other race, including multiracial participants. These demographic categories were evaluated as candidate predictors. Smoking history was obtained by questionnaire; former smokers had smoked at least 100 cigarettes during their lifetime but did not currently smoke. Alcohol intake was assessed from two 24-hour dietary recalls. Hypertension was identified from systolic blood pressure >140 mmHg, diastolic blood pressure >90 mmHg, or current prescribed antihypertensive medication. Diabetes was identified from HbA1c ⩾6.5%, fasting plasma glucose ⩾126 mg/dL (7.0 mmol/L), a reported physician diagnosis outside pregnancy, or insulin use.

Model establishment and validation

Participants were randomly divided into training and testing sets at an approximate 7:3 ratio. Model development used the training set only. Binary logistic least absolute shrinkage and selection operator (LASSO) regression [22] considered 60 candidate predictor columns, with $\alpha = 1$, 100 candidate λ values, and a random seed of 300. Tenfold cross-validation used binomial deviance and lambda.min (the minimum cross-validation error) selected the penalty. LASSO-selected variables entered unadjusted logistic regression; those with $p < 0.1$ entered multivariable logistic regression, and variables with $p < 0.05$ were retained for the final seven-predictor model and nomogram. The fixed predictor set and coefficients estimated in the training set were applied unchanged to the testing set; no predictors or prediction coefficients were re-estimated using the testing data.

Discrimination was assessed by the area under the receiver operating characteristic curve (AUC), and paired AUCs were compared using DeLong tests [23]. The probability cut-off maximizing the Youden index in the training set was applied unchanged to the testing set to calculate sensitivity and specificity. Calibration was assessed graphically and by logistic recalibration with both an intercept and a slope [24]; these assessment coefficients were not used to update the prediction model. Internal bootstrap validation used 1,000 training-set resamples: the coefficients of the final logistic model were re-estimated in each resample and performance was evaluated in that resample and the original training set to estimate mean optimism. This procedure evaluated the final predictor set rather than repeating the preceding variable-selection process.

Four models were compared: M1 included BMI and height; M2 added age and race; M3 added triglycerides, phosphorus, and chloride to M2 (the original seven-predictor model); and M4 included age, race, triglycerides, phosphorus, and chloride, excluding BMI and height. Training-set likelihood-ratio tests assessed nested-model fit. Decision curve analysis (DCA) compared net benefit across probability thresholds with treat-all and treat-none strategies [25]. Training-set curves represent apparent performance; testing-set curves assess performance in the held-out sample. No clinically validated probability threshold for recommending further muscle assessment was established.

Statistical analysis

Analyses were performed using R (version 4.2.1; R Foundation for Statistical Computing, Vienna, Austria). Descriptive summaries and between-group comparisons in Table *1* accounted for NHANES sampling weights, strata (SDMVSTRA), and primary sampling units (SDMVPSU). Counts are unweighted; percentages and medians with 25th and 75th percentiles describe the survey-weighted sample. Model development, logistic regression, bootstrap validation, model comparisons, and DCA were unweighted. Tests were two-sided, with $p < 0.05$ considered statistically significant.

Patient and public involvement

Patients and the public were not involved in the design or conduct of this secondary analysis. Results were not disseminated directly to individual NHANES participants.

Results

Baseline characteristics

The analytic sample included 1,571 adults with NAFLD, comprising 781 men and 790 women, with a survey-weighted median age of 43.0 years (Table *1*). Low appendicular muscle mass was present in 240 participants (unweighted prevalence: 15.28%; survey-weighted prevalence: 14.0%). Compared with participants without low appendicular muscle mass, those with the outcome had higher BMI (35.8 versus 32.9 kg/m²), shorter height (162.0 versus 170.3 cm), and greater waist circumference (114.5 versus 108.4 cm; all $p < 0.001$). Survey-weighted diabetes prevalence was 27% versus 16%, and reported moderate or vigorous recreational activity was 34% versus 54%. Other demographic and laboratory characteristics are summarized in Table *1*.

Characteristic	Total N = 1,571	Without low appendicular muscle mass n = 1,331 (86%)	With low appendicular muscle mass n = 240 (14%)	P value
Age (years)	43.0 (32.1, 51.0)	43.0 (32.0, 51.0)	44.0 (34.0, 53.0)	0.3
Sex				0.5
Male	781 (53%)	665 (53%)	116 (56%)	
Female	790 (47%)	666 (47%)	124 (44%)	
Race/ethnicity				0.001
Mexican American	312 (14%)	223 (12%)	89 (27%)	

Other Hispanic	189 (8.3%)	152 (7.9%)	37 (11%)	
Non-Hispanic White	546 (61%)	473 (62%)	73 (51%)	
Non-Hispanic Black	310 (10%)	293 (11%)	17 (3.9%)	
Non-Hispanic Asian	158 (4.2%)	137 (4.2%)	21 (4.4%)	
Other race including multiracial	56 (2.6%)	53 (2.7%)	3 (1.9%)	
BMI (kg/m²)	33.2 (29.8, 37.2)	32.9 (29.3, 36.6)	35.8 (32.5, 41.3)	<0.001
Height (cm)	169.1 (162.8, 176.3)	170.3 (163.7, 177.1)	162.0 (153.8, 169.8)	<0.001
Waist circumference (cm)	109.0 (101.5, 118.4)	108.4 (101.0, 117.4)	114.5 (105.1, 125.6)	<0.001
Weight (kg)	94.8 (84.8, 107.3)	94.8 (85.5, 107.3)	95.3 (81.4, 107.8)	0.7
Hypertension				0.2
Yes	490 (30%)	406 (30%)	84 (35%)	
No	1,081 (70%)	925 (70%)	156 (65%)	
Systolic pressure (mmHg)	121.0 (112.7, 130.0)	120.7 (112.7, 129.3)	122.0 (112.8, 133.3)	0.089
Diastolic pressure (mmHg)	74.0 (67.7, 80.0)	74.0 (68.0, 80.0)	74.0 (66.7, 80.7)	>0.9
Diabetes				<0.001
Yes	329 (17%)	263 (16%)	66 (27%)	
No	1,242 (83%)	1,068 (84%)	174 (73%)	
Moderate or vigorous recreational activity				<0.001
Yes	774 (51%)	685 (54%)	89 (34%)	
No	797 (49%)	646 (46%)	151 (66%)	
Smoking status				>0.9
Current smoker	343 (21%)	298 (21%)	45 (20%)	
Former smoker	281 (21%)	233 (21%)	48 (22%)	
Never smoker	947 (57%)	800 (57%)	147 (58%)	
White blood cell count (10^3 cells/μL)	7.2 (5.9, 8.5)	7.1 (5.9, 8.4)	7.6 (6.3, 8.9)	0.021
Lymphocyte count (10^3 cells/μL)	2.1 (1.8, 2.6)	2.1 (1.8, 2.5)	2.1 (1.8, 2.6)	0.4
Monocyte percent (%)	7.4 (6.3, 8.8)	7.5 (6.4, 8.9)	7.0 (5.9, 8.4)	0.005
Segmented neutrophils percent (%)	58.8 (52.6, 64.4)	58.5 (52.3, 64.1)	60.9 (55.2, 65.9)	0.012
Eosinophils percent (%)	2.3 (1.6, 3.4)	2.3 (1.5, 3.5)	2.2 (1.6, 3.2)	0.2
Basophils percent (%)	0.7 (0.5, 0.9)	0.7 (0.5, 0.9)	0.7 (0.5, 0.9)	0.7
Red blood cell count (10^6 cells/μL)	4.9 (4.6, 5.2)	4.9 (4.6, 5.2)	4.9 (4.5, 5.3)	0.7
Hemoglobin (g/dL)	14.6 (13.5, 15.5)	14.6 (13.5, 15.5)	14.6 (13.6, 15.6)	0.5

Hematocrit (%)	43.1 (40.3, 45.7)	43.0 (40.2, 45.5)	43.2 (40.5, 46.4)	0.3
Mean cell volume (fL)	88.5 (85.1, 91.3)	88.4 (85.1, 91.3)	88.6 (84.9, 91.3)	0.8
Mean cell hemoglobin (pg)	30.0 (28.6, 31.1)	30.0 (28.6, 31.1)	30.0 (28.3, 31.1)	0.6
Mean cell hemoglobin concentration (g/dL)	33.8 (33.1, 34.4)	33.8 (33.1, 34.4)	33.6 (33.1, 34.3)	0.2
Mean platelet volume (fL)	8.3 (7.8, 9.0)	8.3 (7.8, 9.0)	8.3 (7.7, 8.9)	0.7
Red cell distribution width (%)	13.3 (12.8, 13.9)	13.3 (12.8, 13.9)	13.5 (12.9, 14.1)	0.019
Platelet count (10^3 cells/μL)	241.0 (206.0, 281.0)	238.0 (206.0, 279.0)	256.0 (214.7, 292.0)	0.089
Albumin (g/L)	42.0 (40.0, 44.0)	42.0 (40.0, 44.0)	42.0 (39.0, 44.0)	0.3
Globulin (g/L)	29.0 (26.0, 32.0)	29.0 (26.0, 32.0)	29.0 (27.0, 33.0)	0.1
Total protein (g/L)	71.0 (68.0, 74.0)	71.0 (68.0, 74.0)	72.0 (69.0, 74.9)	0.4
Gamma glutamyl transferase (U/L)	24.0 (17.0, 39.0)	24.0 (17.0, 38.0)	27.0 (19.0, 41.2)	0.12
Aspartate aminotransferase (U/L)	22.0 (19.0, 28.0)	22.0 (18.0, 28.0)	24.0 (19.0, 30.0)	0.048
Alanine aminotransferase (U/L)	25.0 (18.0, 35.0)	25.0 (18.0, 35.0)	26.0 (18.0, 39.0)	0.093
Alkaline phosphatase (U/L)	71.0 (58.0, 84.0)	70.0 (57.0, 83.0)	74.0 (64.0, 91.0)	0.007
Total bilirubin (μmol/L)	10.3 (6.8, 12.0)	10.3 (6.8, 12.0)	10.3 (6.8, 12.0)	0.8
Blood urea nitrogen (mg/dL)	12.0 (10.0, 15.0)	12.0 (10.0, 15.0)	11.0 (9.0, 14.0)	0.15
Creatinine (μmol/L)	72.5 (61.9, 84.3)	73.4 (61.9, 85.8)	68.1 (57.5, 77.8)	<0.001
Creatine phosphokinase (IU/L)	108.0 (77.0, 167.0)	110.0 (78.0, 172.7)	93.0 (65.9, 131.5)	<0.001
Lactate dehydrogenase (U/L)	129.0 (113.0, 147.0)	128.0 (113.0, 147.0)	132.0 (116.0, 151.6)	0.3
Total calcium (mmol/L)	2.3 (2.3, 2.4)	2.3 (2.3, 2.4)	2.3 (2.3, 2.4)	0.6
Iron (μmol/L)	14.3 (10.7, 18.6)	14.1 (10.7, 18.6)	14.5 (10.4, 19.7)	0.8
Phosphorus (mmol/L)	1.2 (1.0, 1.3)	1.2 (1.0, 1.3)	1.1 (1.0, 1.3)	0.6
Uric acid (μmol/L)	345.0 (291.5, 392.6)	339.8 (285.5, 386.6)	356.9 (297.4, 406.3)	0.03
Sodium (mmol/L)	139.0 (138.0, 140.0)	139.0 (138.0, 140.0)	139.0 (138.0, 141.0)	>0.9
Potassium (mmol/L)	4.0 (3.8, 4.2)	4.0 (3.8, 4.2)	4.0 (3.8, 4.3)	>0.9
Chloride (mmol/L)	103.0 (102.0, 105.0)	103.5 (102.0, 105.0)	103.0 (101.0, 105.0)	0.2
	278.0 (275.0,			

Osmolality (mmol/kg)	281.0)	278.0 (275.0, 281.0)	278.0 (275.0, 282.0)	0.5
Direct HDL cholesterol (mg/dL)	45.0 (38.0, 52.0)	45.0 (38.0, 52.0)	43.0 (37.0, 52.2)	0.5
Triglycerides (mmol/L)	1.4 (1.0, 2.1)	1.4 (1.0, 2.1)	1.6 (1.1, 2.0)	0.071
LDL-cholesterol (mmol/L)	3.0 (2.5, 3.6)	3.0 (2.5, 3.6)	3.1 (2.5, 3.6)	0.3
Total cholesterol (mmol/L)	5.0 (4.3, 5.6)	5.0 (4.3, 5.6)	5.0 (4.5, 5.6)	0.3
Total 25-hydroxyvitamin D_2 + D_3 (nmol/L)	60.9 (46.2, 77.0)	61.1 (46.6, 77.2)	56.7 (44.4, 76.0)	0.2
25-hydroxyvitamin D_2 (nmol/L)	1.5 (1.5, 1.5)	1.5 (1.5, 1.5)	1.5 (1.5, 1.5)	0.6
25-hydroxyvitamin D_3 (nmol/L)	57.4 (42.9, 73.6)	58.0 (43.3, 74.6)	51.3 (39.0, 69.7)	0.033

TABLE 1: Characteristics of participants with NAFLD by low appendicular muscle mass status in NHANES 2011-2018.

Values are unweighted n (survey-weighted %) or survey-weighted median (25th, 75th percentile). The percentages in the outcome-group headers are survey-weighted; n = 240/1,571 corresponds to an unweighted prevalence of 15.28%. Summaries and comparisons account for sampling weights, SDMVSTRA, and SDMVPSU. Table 1 uses all 1,571 participants; Table 2 uses the unweighted training set (n = 1,099), so their P values are not directly comparable.

NAFLD, nonalcoholic fatty liver disease; NHANES, National Health and Nutrition Examination Survey; BMI, body mass index; HDL, high-density lipoprotein; LDL, low-density lipoprotein

Performance and validation of the nomogram

Participants were randomly divided into training (n = 1,099) and testing (n = 472) sets, with 168 (15.29%) and 72 (15.25%) outcome events, respectively. LASSO selection is shown in Figure *2*. The selected variables and subsequent unadjusted and multivariable regression results are reported in Table *2*. Seven variables meeting the multivariable selection criterion were retained for the final model: age, race/ethnicity, BMI, height, triglycerides, phosphorus, and chloride (Figure *3*). In the multivariable selection-stage analysis, older age, higher BMI, higher triglycerides, shorter height, and lower phosphorus and chloride were associated with higher odds of the outcome. Non-Hispanic Black participants had lower adjusted odds than non-Hispanic White participants (OR = 0.234, 95% CI: 0.109-0.501); other race/ethnicity comparisons were not statistically significant (Table *2*). As a model-based illustration, the estimated probability for a 45-year-old non-Hispanic White adult with NAFLD, BMI of 30 kg/m², height of 170 cm, triglycerides of 1.5 mmol/L, phosphorus of 1.2 mmol/L, and chloride of 103 mmol/L was approximately 5.65%.

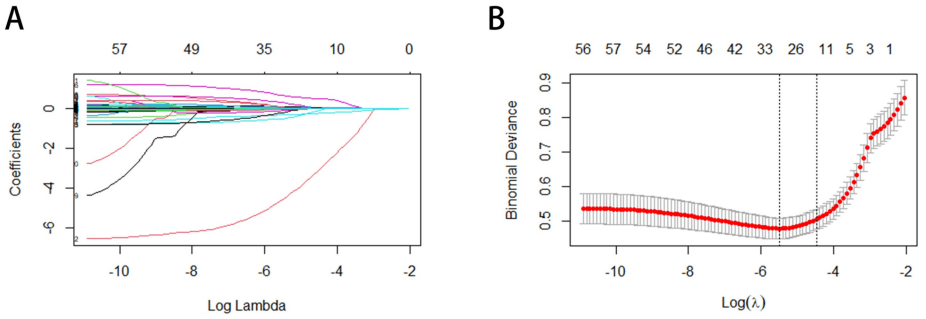


FIGURE 2: Variable selection using LASSO regression.

(A) Regression coefficient paths. (B) Cross-validation curve for the penalty parameter.

LASSO, least absolute shrinkage and selection operator; λ, penalty parameter

Variable	Unadjusted OR (95% CI)	Unadjusted P value	Multivariable selection-stage OR (95% CI)	P value

Race/ethnicity				
Non-Hispanic White	Reference		Reference	
Non-Hispanic Asian	0.795 (0.407-1.553)	0.503	0.726 (0.337-1.567)	0.415
Non-Hispanic Black	0.427 (0.226-0.806)	0.009	0.234 (0.109-0.501)	<0.001
Other race including multiracial	0.502 (0.150-1.685)	0.265	0.463 (0.130-1.646)	0.234
Mexican American	2.642 (1.730-4.036)	<0.001	1.508 (0.905-2.513)	0.115
Other Hispanic	1.958 (1.187-3.228)	0.008	0.991 (0.550-1.785)	0.976
Age (years)	1.028 (1.012-1.044)	<0.001	1.027 (1.008-1.047)	0.005
Sex				
Female	Reference			
Male	0.941 (0.677-1.307)	0.717		
BMI (kg/m²)	1.066 (1.041-1.091)	<0.001	1.103 (1.069-1.139)	<0.001
Height (cm)	0.886 (0.866-0.906)	<0.001	0.885 (0.862-0.908)	<0.001
Segmented neutrophils percent (%)	1.024 (1.004-1.044)	0.016	1.014 (0.990-1.039)	0.265
Eosinophils percent (%)	0.932 (0.848-1.025)	0.146		
Basophils percent (%)	1.153 (0.786-1.691)	0.467		
Red blood cell count (10^6 cells/μL)	1.020 (0.729-1.427)	0.908		
Mean cell hemoglobin (pg)	0.997 (0.935-1.063)	0.929		
Mean platelet volume (fL)	1.110 (0.935-1.318)	0.231		
Direct HDL cholesterol (mg/dL)	1.002 (0.988-1.016)	0.815		
Triglycerides (mmol/L)	1.238 (1.024-1.498)	0.028	1.369 (1.067-1.756)	0.014
Albumin (g/L)	0.984 (0.939-1.031)	0.492		
Globulin (g/L)	0.997 (0.962-1.034)	0.883		
Total protein (g/L)	0.987 (0.951-1.024)	0.493		
Alkaline phosphatase (U/L)	1.010 (1.004-1.016)	0.001	1.006 (0.998-1.014)	0.125
Blood urea nitrogen (mg/dL)	1.008 (0.976-1.041)	0.637		
Creatine phosphokinase (IU/L)	0.999 (0.997-1.000)	0.062	1.001 (1.000-1.002)	0.059
Phosphorus (mmol/L)	0.367 (0.141-0.956)	0.04	0.260 (0.084-0.801)	0.019
Chloride (mmol/L)	0.948 (0.895-1.004)	0.067	0.933 (0.870-0.999)	0.048
Osmolality (mmol/kg)	1.023 (0.987-1.059)	0.211		
Total 25-hydroxyvitamin D_2 + D_3 (nmol/L)	1.001 (0.994-1.008)	0.772		
25-hydroxyvitamin D_3 (nmol/L)	0.994 (0.987-1.002)	0.137		
Diastolic pressure (mmHg)	0.993 (0.981-1.006)	0.297		
Smoking status				
Former/current smoker	Reference			
Never smoker	1.010 (0.722-1.414)	0.953		
Diabetes				
No	Reference		Reference	
Yes	1.595 (1.095-2.324)	0.015	0.890 (0.547-1.449)	0.639

TABLE 2: Unadjusted and multivariable selection-stage logistic regression analyses for low appendicular muscle mass.

Unweighted analyses in the training set (n = 1,099). Each unadjusted model contains one candidate variable. Variables with p < 0.1 entered the multivariable selection-stage model; those with p < 0.05 were retained for the final seven-predictor model. Continuous-variable ORs are per unit shown. Table 1 uses the full, survey-weighted sample.

OR, odds ratio; CI, confidence interval; BMI, body mass index; HDL, high-density lipoprotein

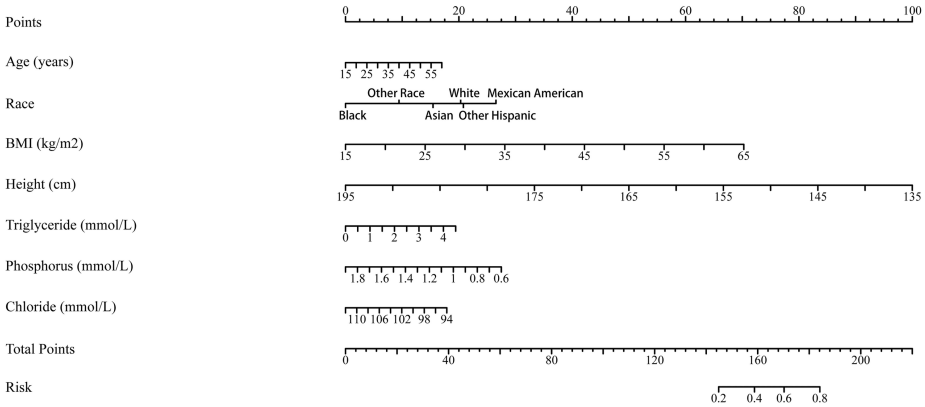


FIGURE 3: Nomogram for identifying prevalent low appendicular muscle mass in patients with NAFLD.

The nomogram estimates the probability of prevalent low appendicular muscle mass, not future muscle loss. The age-axis tick at 15 years is a plotting label; the analytic sample contained only adults aged 18-59 years, and predictions outside that range were not validated. Race labels follow the NHANES demographic categories.

NAFLD, nonalcoholic fatty liver disease; NHANES: National Health and Nutrition Examination Survey; BMI, body mass index

The model AUC was 0.838 (95% CI: 0.808-0.868) in the training set and 0.812 (95% CI: 0.757-0.868) in the held-out testing set (Figure *4*; Appendices, Supplementary Table S1). At the training-derived probability cut-off of 0.1838, sensitivity and specificity were 70.24% and 79.38% in the training set and 65.28% and 78.25% in the testing set. Calibration is shown in Figure *5*. The testing-set logistic recalibration intercept was 0.0010, and the slope was 0.9725. In 1,000 bootstrap resamples, the apparent training AUC of 0.8378 had mean optimism of 0.0121, giving an optimism-corrected AUC of 0.8257 and an optimism-corrected calibration slope of 0.9279.

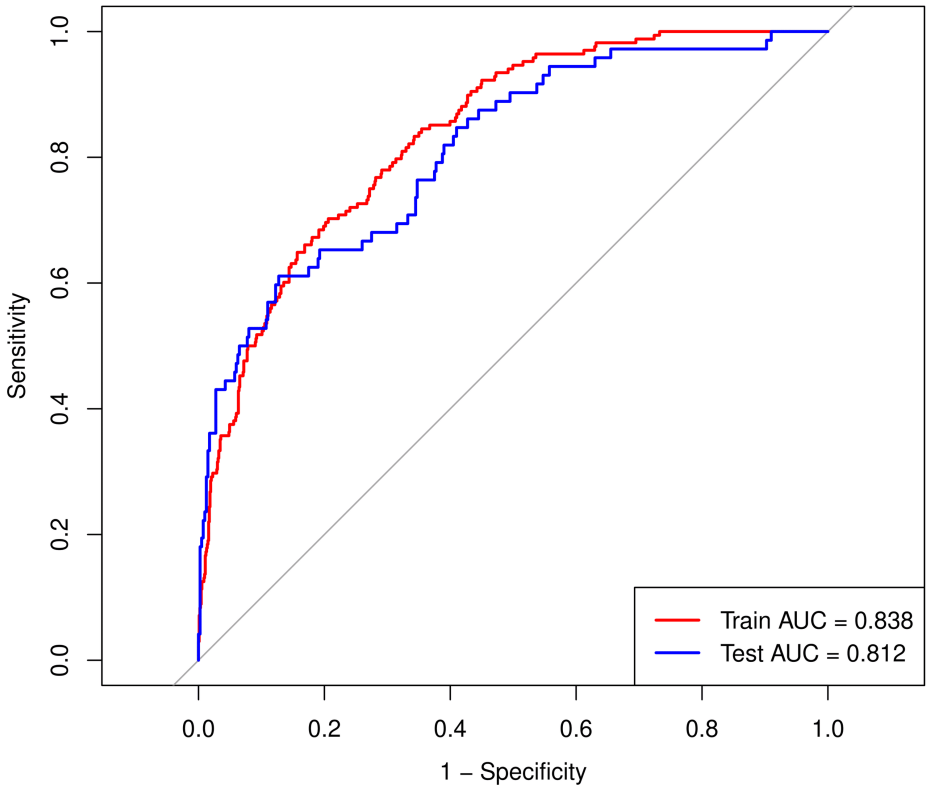


FIGURE 4: Receiver operating characteristic curves for the nomogram.

The red curve represents the training set, AUC = 0.838 (95% CI: 0.808-0.868); the blue curve represents the testing set, AUC = 0.812 (95% CI: 0.757-0.868). The fixed training-set model was applied to the testing set.

AUC, area under the curve; CI, confidence interval

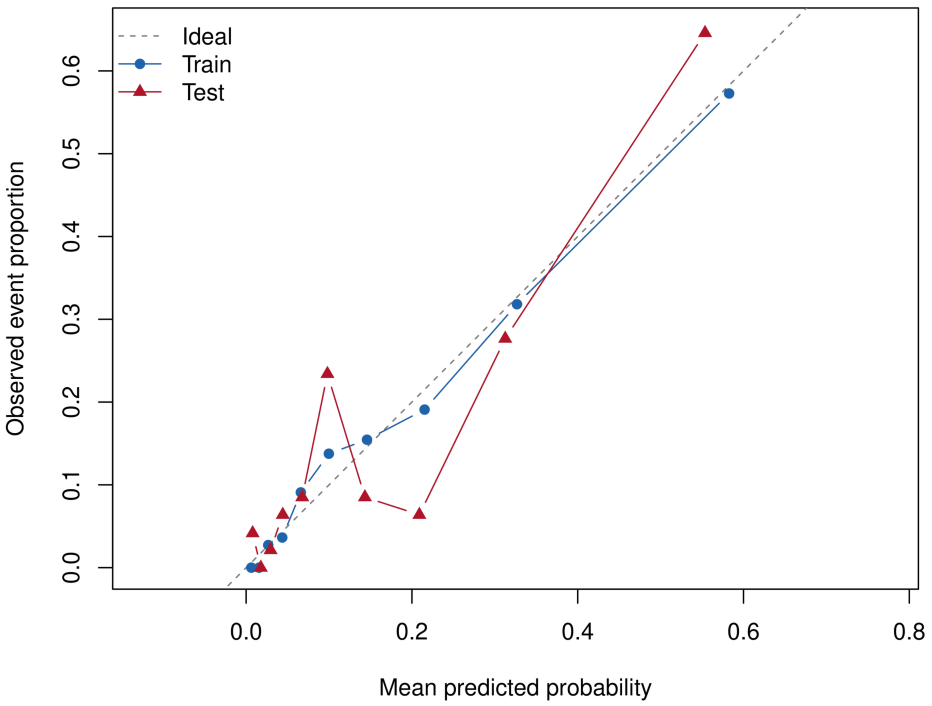


FIGURE 5: Calibration of the seven-predictor model in the training and testing sets.

Blue circles represent the training set and red triangles the testing set; the dashed diagonal represents ideal calibration. Points compare mean predicted probabilities with observed proportions. The testing recalibration intercept (0.0010) and slope (0.9725) were estimated for assessment only; they did not update predictions. Decision curves comparing all four models are shown in the Appendices (Supplementary Figure S2).

Four-model comparisons are presented in the Appendices (Supplementary Tables S1-S4 and Figures S1-S2). Testing-set AUCs were 0.787 for M1, 0.813 for M2, 0.812 for M3, and 0.648 for M4. M3 did not significantly improve testing discrimination over M1 (DeLong p = 0.188) or M2 (p = 0.947). Adding the laboratory variables to M2 increased the training AUC from 0.825 to 0.838 (p = 0.017), without a corresponding testing-set gain. Excluding BMI and height reduced the AUC from 0.8378 to 0.7010 in training and from 0.8123 to 0.6480 in testing (M3 versus M4, both p < 0.001); the testing calibration slope declined from 0.9725 to 0.6424. Testing-set DCA showed relatively higher M3 net benefit at some thresholds, but M3 did not consistently outperform M2 across thresholds. Training-set likelihood-ratio tests assess model fit and do not establish incremental held-out predictive benefit.

Discussion

This study developed and internally validated a seven-predictor nomogram for prevalent low appendicular muscle mass in adults with NAFLD. The held-out testing AUC was 0.812, with a calibration slope of 0.9725. Additional analyses showed that much of its discrimination depended on anthropometric predictors and that adding laboratory variables to age, race/ethnicity, BMI, and height did not improve testing-set discrimination. These findings support further evaluation of the model but do not establish a clinical advantage from the laboratory measurements.

The association between muscle abnormalities and NAFLD has been reported across epidemiological and clinical studies [4,26]. Skeletal muscle participates in metabolic regulation, while aging, systemic inflammation, and insulin resistance can contribute to muscle loss [27]. Sarcopenic obesity illustrates the coexistence of excess adiposity with impaired muscle mass and function [28]. These mechanisms provide context for muscle assessment in NAFLD, but the outcome in this study was limited to a BMI-adjusted muscle-mass measure and did not capture the full clinical syndrome of sarcopenia.

Previous research has reported variation in the prevalence of sarcopenia across race/ethnicity groups and across operational definitions [29]. In this study, non-Hispanic Black participants had lower adjusted odds of low appendicular muscle mass than non-Hispanic White participants (OR = 0.234, 95% CI: 0.109-0.501). The adjusted comparisons for Mexican American, other Hispanic, non-Hispanic Asian, and other race/multiracial participants were not statistically significant (Table *2*). These estimates describe associations with self-reported demographic categories. Differences in body composition, the applicability of muscle-mass cutoffs, and unmeasured social or clinical factors may contribute to the observed

association; it should not be interpreted as evidence of an inherent biological protective effect.

Mathematical coupling is important when interpreting the model. BMI appears in the denominator of ASM/BMI and is also a predictor, while height is structurally related to BMI and body size. Removing BMI and height substantially reduced discrimination, but this comparison does not eliminate dependence on the original BMI-adjusted outcome. FLI also contains BMI and triglycerides, creating a potential dependency between eligibility and model predictors [17]. Although liver-muscle interactions and myosteatosis may contribute to the clinical association [11], these mechanisms were not measured here. Alternative muscle-mass definitions that do not divide by BMI are needed to assess model robustness.

Lower phosphorus and chloride were associated with higher odds of the outcome in the multivariable selection-stage analysis. The phosphorus estimate was imprecise (OR = 0.260, 95% CI: 0.084-0.801), and the chloride interval was close to the null (OR = 0.933, 95% CI: 0.870-0.999). Chloride conductance contributes to skeletal muscle excitability [30], but this does not establish a causal role for serum chloride in the observed association. The results do not support electrolyte correction as an intervention for low appendicular muscle mass.

Strengths include DXA-based muscle assessment, application of fixed training-set models to a held-out testing set, and comparisons with simpler models. However, the cohort was restricted to adults aged 18-59 years with complete DXA and covariate data. Of 4,907 eligible participants with NAFLD, 2,485 lacked DXA data, which may introduce selection bias. Survey-weighted descriptive statistics do not make the unweighted model performance representative of the overall US population. Applicability to older adults and other populations remains uncertain.

Several further limitations affect interpretation. The FNIH cutoffs were developed in older populations [21], and their clinical applicability in this younger sample requires evaluation. Strength and physical performance were not assessed as part of the outcome [6,7]. NAFLD definitions differed between FLI-based and CAP-based survey cycles [17,18], and the liver stiffness exclusion was applied only in 2017-2018. These differences may introduce heterogeneity; no fibrosis-stage-specific analysis was performed. The cross-sectional design cannot establish temporal or causal relationships, predict future muscle loss, or assess the effect of model-guided intervention.

The laboratory variables improved apparent training performance but did not improve testing discrimination over the model containing age, race/ethnicity, BMI, and height. The comparison with BMI and height alone also added age and race/ethnicity, so any difference cannot be attributed solely to laboratory measurements. DCA showed threshold-dependent differences, without a clinically validated threshold for further muscle assessment. The Youden cut-off is a statistical operating point rather than an established referral threshold. Prospective studies should compare clinical pathways, testing burden, and patient outcomes before implementation.

The training set contained 168 outcome events relative to 60 candidate predictor columns. LASSO followed by two P-value selection steps may produce unstable selection and overfitting. Bootstrap validation refitted the final predictor set without repeating the preceding selection pipeline and therefore did not capture all selection uncertainty. A single split and bootstrap assessment cannot substitute for external validation.

If evaluated prospectively, the proposed model could be studied as an aid for clinicians deciding whether patients with existing anthropometric and laboratory results warrant further muscle assessment. It should not replace direct measurements or a clinical sarcopenia assessment. External validation using consistent liver-disease criteria and alternative muscle-mass outcomes is needed to determine whether this use offers benefit beyond simpler models.

Conclusions

The seven-predictor nomogram showed internal discrimination for prevalent low appendicular muscle mass in selected adults aged 18-59 years with NAFLD. Performance depended substantially on anthropometry and should be interpreted in light of the BMI-adjusted outcome and multistage selection. Adding laboratory variables did not improve held-out testing discrimination over age, race/ethnicity, BMI, and height. External validation and prospective evaluation of clinical usefulness are required before implementation.

Appendices

Performance metric	Training set	Testing set
Sample size	1,099	472
Low appendicular muscle mass cases	168	72
AUC	0.838	0.812
95% CI	0.808-0.868	0.757-0.868
Probability cut-off	0.1838	0.1838
Sensitivity (%)	70.24	65.28
Specificity (%)	79.38	78.25

TABLE 3: Supplementary Table S1: Discriminative performance of the seven-predictor model in the training and testing sets.

The probability cut-off was determined by maximizing the Youden index in the training set and was applied unchanged to the testing set.

AUC, area under the receiver operating characteristic curve; CI, confidence interval

Model	Dataset	N	Events	AUC (95% CI)	Calibration slope	Calibration intercept*
M1	Train	1,099	168	0.783 (0.748-0.818)	1	0
M1	Test	472	72	0.787 (0.731-0.843)	1.029	0.063
M2	Train	1,099	168	0.825 (0.794-0.856)	1	0
M2	Test	472	72	0.813 (0.758-0.868)	1.012	0.043
M3	Train	1,099	168	0.838 (0.808-0.868)	1	0
M3	Test	472	72	0.812 (0.757-0.868)	0.972	0.001
M4	Train	1,099	168	0.701 (0.660-0.742)	1	0
M4	Test	472	72	0.648 (0.579-0.717)	0.642	-0.584

TABLE 4: Supplementary Table S2: Discrimination and calibration of the four prediction models.

M1 includes body mass index (BMI) and height; M2 adds age and race/ethnicity to M1; M3 adds triglycerides, phosphorus, and chloride to M2; and M4 includes the M3 predictors except BMI and height. Counts and analyses are unweighted; events are cases of low appendicular muscle mass. The calibration intercept marked with an asterisk was estimated from a logistic recalibration model containing both an intercept and a slope; it is not the calibration-in-the-large estimate. Apparent training-set calibration slopes equal 1 by construction. Testing-set calibration estimates were used for assessment only and did not update the predictions.

Train, training set; Test, testing set; AUC, area under the receiver operating characteristic curve; CI, confidence interval

Dataset	Comparison	ΔAUC (M3 - comparator)	DeLong P
Train	M3 vs M1	0.0548	<0.001
Train	M3 vs M2	0.0125	0.017
Train	M3 vs M4	0.1368	<0.001
Test	M3 vs M1	0.0253	0.188
Test	M3 vs M2	-0.0007	0.947
Test	M3 vs M4	0.1643	<0.001

TABLE 5: Supplementary Table S3: Paired comparisons of the areas under the receiver operating characteristic curve.

M1 includes body mass index (BMI) and height; M2 adds age and race/ethnicity to M1; M3 adds triglycerides, phosphorus, and chloride to M2; and M4 includes the M3 predictors except BMI and height. Comparisons use paired DeLong tests in the same participants. ΔAUC is the AUC of M3 minus the AUC of the comparator model.

Train, training set; Test, testing set; AUC, area under the receiver operating characteristic curve

Comparison	χ^2	df	P
M3 vs M1	64.877	9	<0.001
M3 vs M2	17.888	3	<0.001
M3 vs M4	164.654	2	<0.001

TABLE 6: Supplementary Table S4: Likelihood-ratio tests comparing models in the training set.

M1 includes body mass index (BMI) and height; M2 adds age and race/ethnicity to M1; M3 adds triglycerides, phosphorus, and chloride to M2; and M4 includes the M3 predictors except BMI and height. The likelihood-ratio tests compare nested models using training-set data and assess model fit rather than held-out predictive benefit.

χ^2, chi-square statistic; df, degrees of freedom

Threshold	M1	M2	M3	M4
0.05	0.1204	0.1167	0.1171	0.1028
0.1	0.0742	0.0807	0.0815	0.0669
0.15	0.0543	0.0646	0.0609	0.0313
0.2	0.0461	0.0477	0.0556	0.0201
0.25	0.0325	0.0438	0.0551	0.0028
0.3	0.0409	0.0421	0.0469	0.0042
0.4	0.0374	0.0572	0.0487	-0.012

TABLE 7: Supplementary Table S5: Decision-curve net benefit of the four models in the testing set.

M1 includes body mass index (BMI) and height; M2 adds age and race/ethnicity to M1; M3 adds triglycerides, phosphorus, and chloride to M2; and M4 includes the M3 predictors except BMI and height. Net benefit is reported at selected threshold probabilities in the testing set; these are representative values from the underlying analysis, not the complete set of 99 thresholds. The clinical relevance of these thresholds has not been independently established, and M3 does not uniformly outperform M2. These model comparisons do not eliminate mathematical coupling between the BMI-adjusted outcome and anthropometric predictors.

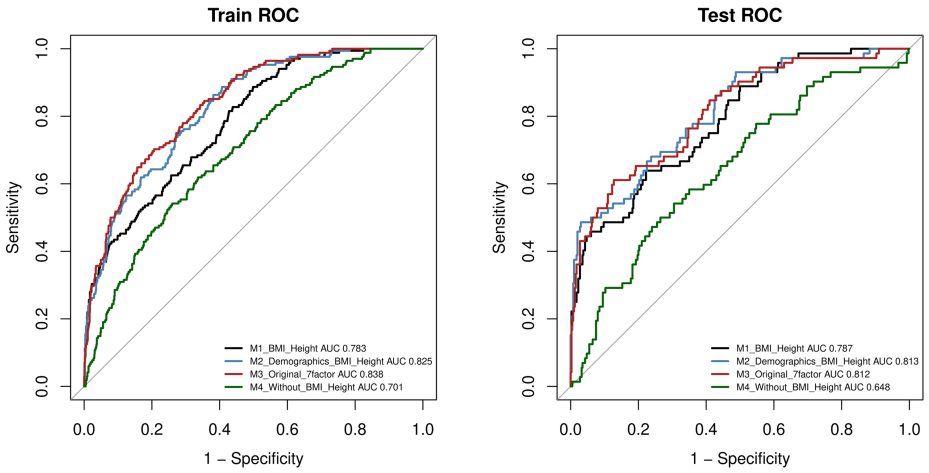


FIGURE 6: Supplementary Figure S1: Receiver operating characteristic curves for the four prediction models.

Training and testing curves correspond to M1-M4 as defined in the Appendices (Supplementary Table S2). All predictions in the testing set were obtained using models developed on the training set.

AUC, area under the receiver operating characteristic curve

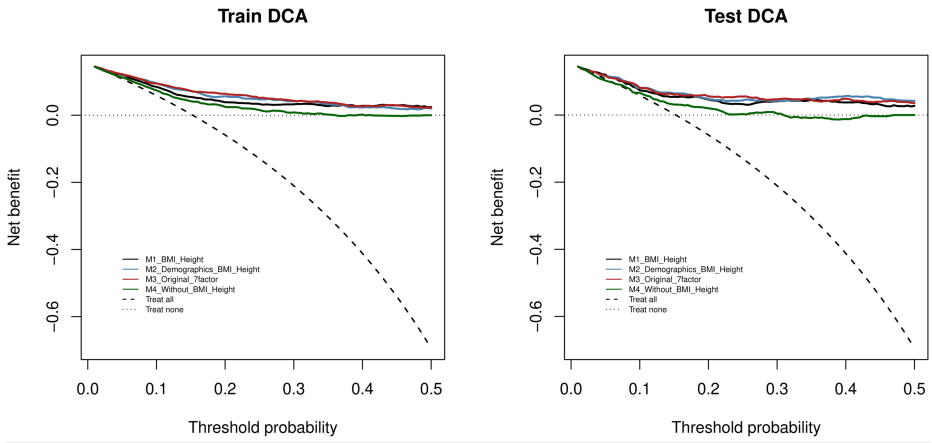


FIGURE 7: Supplementary Figure S2: Decision curves for the four prediction models.

Training-set curves show apparent net benefit; testing-set curves show held-out performance. The treat-all and treat-none strategies are included. Net benefit varies with threshold probability, and no clinically validated referral threshold has been established. Model definitions and selected testing-set net benefit values are provided in the Appendices (Supplementary Table S5).

Section or topic	Item	D or E	Checklist item	Page(s)/Location in the Cureus article
Title	1	D;E	Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	p. 1; Title
Abstract	2	D;E	See TRIPOD+AI for Abstracts checklist	p. 2; Abstract (partial)
Background	3a	D;E	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models	p. 3; Introduction (partial)
Background	3b	D;E	Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users (e.g., healthcare professionals, patients, public)	p. 3; Introduction
Background	3c	D;E	Describe any known health inequalities between sociodemographic groups	p. 6; Discussion (partial)
Objectives	4	D;E	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	p. 3; Introduction
Data	5a	D;E	Describe the sources of data separately for the development and evaluation datasets (e.g., randomised trial, cohort, routine care or registry data), the rationale for using these data, and representativeness of the data	pp. 3, 6-7; Study population; Discussion
Data	5b	D;E	Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	p. 3; Study population
Participants	6a	D;E	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including the number and location of centres	p. 3; Study population (partial)
Participants	6b	D;E	Describe the eligibility criteria for study participants	p. 3; Study population
Participants	6c	D;E	Give details of any treatments received, and how they were handled during model development or evaluation, if relevant	N/A
Data preparation	7	D;E	Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	pp. 3-4; Study population; Variable estimation (partial)

Outcome	8a	D;E	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups	p. 4; Measurements and definition of low appendicular muscle mass (partial)
Outcome	8b	D;E	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	N/A
Outcome	8c	D;E	Report any actions to blind assessment of the outcome to be predicted	NR
Predictors	9a	D	Describe the choice of initial predictors (e.g., literature, previous models, all available predictors) and any pre-selection of predictors before model building	pp. 4, 12; Variable estimation; Model establishment and validation (partial)
Predictors	9b	D;E	Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors for the outcome and other predictors)	pp. 4, 12; Variable estimation; Table 2 (partial)
Predictors	9c	D;E	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	N/A
Sample size	10	D;E	Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question. Include details of any sample size calculation	pp. 3-5, 7; Study population; Results; Discussion (partial)
Missing data	11	D;E	Describe how missing data were handled. Provide reasons for omitting any data	pp. 3, 13; Study population; Figure 1
Analytical methods	12a	D	Describe how the data were used (e.g., for development and evaluation of model performance) in the analysis, including whether the data were partitioned, considering any sample size requirements	pp. 4-5; Model establishment and validation; Results (partial)
Analytical methods	12b	D	Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation).	pp. 4, 12; Variable estimation; Table 2 (partial)
Analytical methods	12c	D	Specify the type of model, rationale, all model-building steps, including any hyperparameter tuning, and method for internal validation	p. 4; Model establishment and validation
Analytical methods	12d	D;E	Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters (e.g., hospitals, countries). See TRIPOD-Cluster for additional considerations	NR
Analytical methods	12e	D;E	Specify all measures and plots used (and their rationale) to evaluate model performance (e.g., discrimination, calibration, clinical utility) and, if relevant, to compare multiple models	pp. 4-5, 16-17; Model establishment and validation; Results; Figures 4-7; Tables 3-7
Analytical methods	12f	E	Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	N/A
Analytical methods	12g	E	For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)	p. 4; Model establishment and validation (partial)
Class imbalance	13	D;E	If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions	NR
Fairness	14	D;E	Describe any approaches that were used to address model fairness and their rationale	NR
Model output	15	D	Specify the output of the prediction model (e.g., probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified	pp. 4-5; Model establishment and validation; Results
Training versus evaluation	16	D;E	Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors	pp. 4-5; Model establishment and validation; Results

Ethical approval	17	D;E	Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent	pp. 3, 7-8; Study population; Ethics approval and consent to participate
Funding	18a	D;E	Give the source of funding and the role of the funders for the present study	p. 8; Funding
Conflicts of interest	18b	D;E	Declare any conflicts of interest and financial disclosures for all authors	p. 8; Competing interests
Protocol	18c	D;E	Indicate where the study protocol can be accessed or state that a protocol was not prepared	NR
Registration	18d	D;E	Provide registration information for the study, including register name and registration number, or state that the study was not registered	NR
Data sharing	18e	D;E	Provide details of the availability of the study data	p. 8; Availability of data and materials
Code sharing	18f	D;E	Provide details of the availability of the analytical code	NR
Patient & Public Involvement	19	D;E	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement.	p. 5; Patient and public involvement
Participants	20a	D;E	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	pp. 3, 5, 13; Study population; Baseline characteristics; Figure 1
Participants	20b	D;E	Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data. A table may be helpful. Report any differences across key demographic groups.	pp. 3, 5, 10-11; Study population; Baseline characteristics; Table 1 (partial)
Participants	20c	E	For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome).	p. 5; Results; Table 3 (outcome events only)
Model development	21	D;E	Specify the number of participants and outcome events in each analysis (e.g., for model development, hyperparameter tuning, model evaluation)	pp. 4-5; Model establishment and validation; Results; Tables 3 and 4
Model specification	22	D	Provide details of the full prediction model (e.g., formula, code, object, application programming interface) to allow predictions in new individuals and to enable third-party evaluation and implementation, including any restrictions to access or re-use (e.g., freely available, proprietary)	p. 15; Figure 3; full model NR
Model performance	23a	D;E	Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation.	pp. 5-6, 16-17; Results; Figures 4-7; Tables 3-7 (partial)
Model performance	23b	D;E	If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details.	N/A
Model updating	24	E	Report the results from any model updating, including the updated model and subsequent performance	N/A
Interpretation	25	D;E	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	pp. 6-7; Discussion (partial)
Limitations	26	D;E	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalizability	pp. 6-7; Discussion
Usability of the model in the context	27a	D	Describe how poor quality or unavailable input data (e.g., predictor values) should be	NR

of current care			assessed and handled when implementing the prediction model	
Usability of the model in the context of current care	27b	D	Specify whether users will be required to interact in the handling of the input data or use of the model, and what level of expertise is required of users	pp. 3, 7, 15; Introduction; Discussion; Figure 3 (partial)
Usability of the model in the context of current care	27c	D;E	Discuss any next steps for future research, with a specific view to applicability and generalizability of the model	p. 7; Discussion

TABLE 8: Supplementary Table S6: TRIPOD+AI reporting checklist.

D, development; E, evaluation; NR, not reported; N/A, not applicable; partial, partially reported

Notes: 1. D = items relevant only to development; E = items relevant only to evaluation; D;E = items applicable to both. 2. Item 12c applies separately for all model-building approaches. 3. TRIPOD-Cluster provides additional reporting recommendations for studies accounting for clustering or heterogeneity in model performance 4. Item 18f concerns analysis code, including data cleaning, feature engineering, model building, and evaluation. 5. Item 22 concerns code to implement the model for a new individual.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Yujiang Pang, Wenluo Zhang, Zhenhua Hu, Jie Zhou, Zuorui Zhu, Mengni Jing, Yifan Ling

Acquisition, analysis, or interpretation of data: Yujiang Pang, Wenluo Zhang, Zhenhua Hu, Jie Zhou, Yifan Ling

Drafting of the manuscript: Yujiang Pang, Wenluo Zhang, Yifan Ling

Critical review of the manuscript for important intellectual content: Zhenhua Hu, Jie Zhou, Zuorui Zhu, Mengni Jing

Supervision: Zhenhua Hu, Jie Zhou

Disclosures

Human subjects: Informed consent for treatment and open access publication was obtained or waived by all participants in this study. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

Yujiang Pang and Wenluo Zhang contributed equally to this work.

References

1. Le MH, Yeo YH, Li X, et al.: 2019 global NAFLD prevalence: a systematic review and meta-analysis . Clin Gastroenterol Hepatol. 2022, 20:2809-17.e28. 10.1016/j.cgh.2021.12.002
2. Kuchay MS, Martínez-Montoro JI, Kaur P, Fernández-García JC, Ramos-Molina B: Non-alcoholic fatty liver disease-related fibrosis and sarcopenia: an altered liver-muscle crosstalk leading to increased mortality risk. Ageing Res Rev. 2022, 80:101696. 10.1016/j.arr.2022.101696
3. Byrne CD, Targher G: NAFLD: a multisystem disease . J Hepatol. 2015, 62:S47-64. 10.1016/j.jhep.2014.12.012

4. Joo SK, Kim W: Interaction between sarcopenia and nonalcoholic fatty liver disease . Clin Mol Hepatol. 2023, 29:S68-78. 10.3350/cmh.2022.0358
5. European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO): EASL-EASD-EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). J Hepatol. 2024, 81:492-542. 10.1016/j.jhep.2024.04.031
6. Cruz-Jentoft AJ, Bahat G, Bauer J, et al.: Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing. 2019, 48:16-31. 10.1093/ageing/afy169
7. Kirk B, Cawthon PM, Arai H, et al.: The conceptual definition of sarcopenia: Delphi consensus from the Global Leadership Initiative in Sarcopenia (GLIS). Age Ageing. 2024, 53:afae052. 10.1093/ageing/afae052
8. Sayer AA, Cruz-Jentoft A: Sarcopenia definition, diagnosis and treatment: consensus is growing . Age Ageing. 2022, 51:afac220. 10.1093/ageing/afac220
9. Harring M, Golabi P, Paik JM, et al.: Sarcopenia among patients with nonalcoholic fatty liver disease (NAFLD) is associated with advanced fibrosis. Clin Gastroenterol Hepatol. 2023, 21:2876-88.e5. 10.1016/j.cgh.2023.02.013
10. Llanos P, Palomero J: Reactive oxygen and nitrogen species (RONS) and cytokines-myokines involved in glucose uptake and insulin resistance in skeletal muscle. Cells. 2022, 11:4008. 10.3390/cells11244008
11. Marjot T, Armstrong MJ, Stine JG: Skeletal muscle and MASLD: mechanistic and clinical insights . Hepatol Commun. 2025, 9:0711. 10.1097/HC9.0000000000000711
12. Iwaki M, Kobayashi T, Nogami A, Saito S, Nakajima A, Yoneda M: Impact of sarcopenia on non-alcoholic fatty liver disease. Nutrients. 2023, 15:891. 10.3390/nu15040891
13. Hong HC, Hwang SY, Choi HY, et al.: Relationship between sarcopenia and nonalcoholic fatty liver disease: the Korean Sarcopenic Obesity Study. Hepatology. 2014, 59:1772-8. 10.1002/hep.26716
14. Koo BK, Kim D, Joo SK, et al.: Sarcopenia is an independent risk factor for non-alcoholic steatohepatitis and significant fibrosis. J Hepatol. 2017, 66:123-31. 10.1016/j.jhep.2016.08.019
15. Golabi P, Gerber L, Paik JM, Deshpande R, de Avila L, Younossi ZM: Contribution of sarcopenia and physical inactivity to mortality in people with non-alcoholic fatty liver disease. JHEP Rep. 2020, 2:100171. 10.1016/j.jhepr.2020.100171
16. Centers for Disease Control and Prevention, National Center for Health Statistics: about NHANES . (2024). Accessed: October 21, 2026: https://www.cdc.gov/nchs/nhanes/about/.
17. Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, Castiglione A, Tiribelli C: The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterol. 2006, 6:33. 10.1186/1471-230X-6-33
18. 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, 66:1022-30. 10.1016/j.jhep.2016.12.022
19. Eddowes PJ, Sasso M, Allison M, et al.: Accuracy of FibroScan controlled attenuation parameter and liver stiffness measurement in assessing steatosis and fibrosis in patients with nonalcoholic fatty liver disease. Gastroenterology. 2019, 156:1717-30. 10.1053/j.gastro.2019.01.042
20. Centers for Disease Control and Prevention, National Center for Health Statistics: ethics review board approval. (2026). Accessed: October 21, 2026: https://www.cdc.gov/nchs/nhanes/about/erb.html.
21. Cawthon PM, Peters KW, Shardell MD, et al.: Cutpoints for low appendicular lean mass that identify older adults with clinically significant weakness. J Gerontol A Biol Sci Med Sci. 2014, 69:567-75. 10.1093/gerona/glu023
22. Tibshirani R: Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol. 1996, 58:267-88. 10.1111/j.2517-6161.1996.tb02080.x
23. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988, 44:837-45.
24. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW: Calibration: the Achilles heel of predictive analytics. BMC Med. 2019, 17:230. 10.1186/s12916-019-1466-7
25. Vickers AJ, Elkin EB: Decision curve analysis: a novel method for evaluating prediction models . Med Decis Making. 2006, 26:565-74. 10.1177/0272989X06295361
26. Lee YH, Jung KS, Kim SU, et al.: Sarcopaenia is associated with NAFLD independently of obesity and insulin resistance: nationwide surveys (KNHANES 2008-2011). J Hepatol. 2015, 63:486-93. 10.1016/j.jhep.2015.02.051
27. Wiedmer P, Jung T, Castro JP, Pomatto LC, Sun PY, Davies KJ, Grune T: Sarcopenia - molecular mechanisms and open questions. Ageing Res Rev. 2021, 65:101200. 10.1016/j.arr.2020.101200
28. Prado CM, Batsis JA, Donini LM, Gonzalez MC, Siervo M: Sarcopenic obesity in older adults: a clinical overview. Nat Rev Endocrinol. 2024, 20:261-77. 10.1038/s41574-023-00943-z
29. Batsis JA, Barre LK, Mackenzie TA, Pratt SI, Lopez-Jimenez F, Bartels SJ: Variation in the prevalence of sarcopenia and sarcopenic obesity in older adults associated with different research definitions: dual-energy X-ray absorptiometry data from the National Health and Nutrition Examination Survey 1999-2004. J Am Geriatr Soc. 2013, 61:974-80. 10.1111/jgs.12260
30. Renaud JM, Ørtenblad N, McKenna MJ, Overgaard K: Exercise and fatigue: integrating the role of K(+), Na(+) and Cl(-) in the regulation of sarcolemmal excitability of skeletal muscle. Eur J Appl Physiol. 2023, 123:2345-78. 10.1007/s00421-023-05270-9