Cureus
Part of SPRINGER NATURE

Open Access Original Article

Development and Internal Validation of Machine-Learning Survival Models for Predicting Cardiovascular Mortality in Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD)

Review began 11/03/2026
Review ended 11/18/2026
Published 11/25/2026

© **Copyright** 2026
Njei 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.120258

Basile Njei [1], Guy Nguefang [2], Solomon Gyabaah [3], Sarpong Boateng [4], Yussif Issaka [4], Yazan Al-Ajlouni [5], Gordon M. Amponsah [3, 6], Ridwan A. Lawal [7], Ulrick Sidney Kanmounye [8, 9]

1. Medicine, Yale School of Medicine, New Haven, USA 2. Internal Medicine, Texas Tech University Health Sciences Center, Odessa, USA 3. Internal Medicine, Komfo Anokye Teaching Hospital, Kumasi, GHA 4. Medicine, Yale New Haven Health Bridgeport Hospital, Bridgeport, USA 5. Medicine, New York Medical College, New York City, USA 6. Physiology, Kwame Nkrumah Univeristy of Science and Technology, Kumasi, GHA 7. Internal Medicine, University of Texas Rio Grande Valley School of Medicine, Edinburg, USA 8. Neurosurgery, Bel Campus University of Technology, Kinshasa, COD 9. Global Neurosurgery Initiative, Harvard Medical School, Boston, USA

Corresponding author: Solomon Gyabaah, solomongyabaah53@gmail.com

Abstract

Background

Metabolic dysfunction-associated steatotic liver disease (MASLD) is highly prevalent and linked to increased cardiovascular mortality, yet liver-focused risk scores perform poorly for cardiovascular prediction. We aimed to develop and internally validate machine learning-based survival models for long-term cardiovascular mortality in MASLD.

Methods

We analysed data from 3,685 adults with MASLD from the Third National Health and Nutrition Examination Survey (NHANES III), linked to the National Death Index public-use Linked Mortality File. Cardiovascular mortality was defined as death with heart disease recorded as the underlying cause. We developed penalised Cox regression and machine-learning models to predict cardiovascular mortality. Binary classification models were evaluated using a 10-year prediction horizon. Model performance was assessed using apparent performance, stratified five-fold cross-validation, and bootstrap optimism correction based on 200 resamples. Competing risks were evaluated using the Aalen-Johansen cumulative incidence function.

Results

Among 3,685 participants, 504 (13.7%) cardiovascular deaths occurred over a median follow-up of 310 months. In five-fold cross-validation, XGBoost-Cox demonstrated the highest C-index (0.789), while bootstrap validation showed the highest optimism-corrected C-index for XGBoost-Cox (0.806; 95% CI: 0.789-0.820), followed by random forest (0.804; 95% CI: 0.785-0.825). Age and hypertension were the most influential predictors. Models demonstrated generally satisfactory calibration, while decision curve analysis suggested clinical net benefit across threshold probabilities of 5%-30%. Risk stratification identified distinct gradients in cardiovascular mortality, which persisted after accounting for competing risks.

Conclusion

Machine learning models showed strong performance and may improve cardiovascular risk stratification in MASLD.

Categories: Cardiology, Gastroenterology
Keywords: artificial intelligence, cardiovascular mortality, internal validation, machine learning, metabolic dysfunction-associated steatotic liver disease (masld)

Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD) is one of the most common chronic liver conditions worldwide [1,2]. MASLD is a multisystem disease with significant cardiovascular consequences [3]. Lean individuals with MASLD are at an increased risk of cardiovascular mortality [4]. Identifying patients with MASLD who are at heightened risk for cardiovascular mortality is therefore essential for guiding preventive care, early intervention, and risk-adapted management strategies.

How to cite this article
Njei B, Nguefang G, Gyabaah S, et al. (November 25, 2026) Development and Internal Validation of Machine-Learning Survival Models for Predicting Cardiovascular Mortality in Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD). Cureus 18(11): e120258. DOI 10.7759/cureus.120258

Despite this urgent need, current risk-stratification methods remain limited. Widely used liver fibrosis scores such as the Fibrosis-4 (FIB-4) Index and the Non-alcoholic Fatty Liver Disease (NAFLD) Fibrosis Score (NFS) provide valuable prognostic information for hepatic outcomes but perform poorly when applied to cardiovascular risk prediction [5,6]. Traditional statistical models offer baseline prognostication but often fail to capture the nonlinear relationships and complex interactions characteristic of cardiometabolic disease [7]. As a result, clinicians lack accurate, MASLD-specific tools for predicting long-term cardiovascular mortality [8,9].

Traditional prognostic tools (e.g., regression-based scores, fibrosis indices) provide only modest discrimination and calibration for long-term outcomes and may fail to capture complex dependencies among metabolic, hematologic, and demographic predictors [10]. Survival-focused machine-learning methods such as random survival forests, gradient-boosted survival models, and penalised Cox frameworks directly address censored time-to-event data and can learn flexible hazard structures without relying on proportional hazards assumptions, improving discrimination (time-dependent area under the curve (AUC), C-index) and calibration (integrated Brier score) in comparative evaluations [10]. Gradient-boosted survival models can outperform other approaches for all-cause mortality prediction. Survival models are generally preferable when predicting cardiovascular death over a follow-up period because they account for both the occurrence and timing of death, variable follow-up durations, and right-censoring in the assessment of cardiovascular mortality.

Alongside predictive performance, transparency is essential for clinical acceptance and deployment. Explainable AI methods provide human-interpretable rationales for model outputs. In resource-limited clinical settings, XAI has demonstrated value for elucidating nonlinear relationships, quantifying feature importance, and clarifying patient-level decision pathways (e.g., SHapley Additive exPlanations (SHAP), Individual Conditional Expectation (ICE) plots, and breakdown/force plots), enabling targeted interventions and building trust in machine learning-assisted care. This approach has been applied to infectious disease prognostics, where machine learning combined with XAI improved discrimination and interpretability and highlighted behavioural and clinical drivers of adverse outcomes [11].

In this study, we sought to develop and internally validate multiple survival machine-learning models for predicting cardiovascular mortality among adults with MASLD. Our goal was to determine whether machine-learning survival models can improve cardiovascular risk prediction in MASLD and provide a foundation for future externally validated tools that support personalised, evidence-based clinical decision-making.

Materials And Methods

Study design, data source, participants, setting, and reporting

We performed a retrospective cohort study using data from the Third National Health and Nutrition Examination Survey (NHANES III, 1988-1994), a nationally representative survey of the civilian, non-institutionalised U.S. population conducted by the National Center for Health Statistics (NCHS) [12]. The survey employed a multistage, stratified probability sampling methodology and incorporated standardised interviews, physical examinations, and laboratory investigations. Older adults and racial/ethnic minority populations were oversampled to improve the representation of these groups. Participants were followed from their initial examination date until death or the end of follow-up on December 31, 2019, whichever occurred first. Mortality follow-up was ascertained through the publicly available linked mortality data.

We used data from the NHANES III, conducted between 1988 and 1994, linked to the National Death Index public-use Linked Mortality File. Of the 33,994 survey participants, 3,685 adults with MASLD and available mortality follow-up were included in the present analysis.

MASLD was defined as the presence of hepatic steatosis (identified with ultrasound) and at least one of five cardiometabolic risks, including 1) body mass index (BMI) ⩾ 25 kg/m² (23 kg/m² in Asians) or waist circumference > 94 cm in men and 80 cm in women; 2) Fasting serum glucose ⩾ 5.6 mmol/L (100 mg/dl) or two-hour post-load glucose levels ⩾ 7.8 mmol/L (⩾ 140 mg/dl) or HbA1c ⩾ 5.7% (39 mmol/L) or diagnosed with type 2 diabetes or treatment for type 2 diabetes; 3) Blood pressure ⩾ 130/85 mmHg or use of antihypertensive medication; 4) Plasma triglycerides ⩾ 1.70 mmol/L (150 mg/dl) or lipid lowering treatment; 5) Plasma high-density lipoprotein (HDL)-cholesterol ⩽ 1.0 mmol/L (40 mg/dl) in men and ⩽ 1.3 mmol/L (50 mg/dl) in women or lipid lowering treatment.

We compared performance across five models, including penalised (1) Cox proportional hazards, (2) random forest survival, (3) gradient boosting survival, (4) XGBoost survival, and (5) logistic regression. Participants were included if they met criteria for MASLD, defined using established metabolic and hepatic biomarker thresholds [13,14]. Individuals with missing outcome data or incomplete covariate information after imputation were excluded. Cardiovascular mortality, determined through linked National Death Index data, served as the primary outcome.

The findings of this study are reported in accordance with the Transparent Reporting of a multivariable

prediction model for Individual Prognosis Or Diagnosis-Artificial Intelligence (TRIPOD+AI) [15].

Outcome definition and follow-up

The date of the NHANES interview was considered the time origin, and follow-up duration was measured in months. Participants who remained alive at the end of the available follow-up period were administratively censored. Deaths from causes other than cardiovascular disease were considered competing events. For the cause-specific Cox and XGBoost-Cox models, these deaths were censored at the time of death, allowing estimation of cause-specific hazards. Their impact on the absolute probability of cardiovascular death was additionally evaluated using the Aalen-Johansen cumulative incidence function.

Outcome ascertainment

Cardiovascular mortality, determined through linked National Death Index data, served as the primary outcome. Follow-up time began from the date of NHANES III baseline examination to the date of cardiovascular death or censoring.

Predictors and missing data analysis

Based on the clinical relevance to metabolic and cardiovascular risk, 23 candidate predictors were selected. These included demographic factors (age, sex), metabolic comorbidities (diabetes, hypertension, dyslipidemia), anthropometrics (BMI, waist circumference), laboratory biomarkers (creatinine, fasting glucose, lipid profile, bilirubin), and fibrosis-related indices (FIB-4, NFS). Missing data were addressed using multivariable imputation techniques.

Missing data and preprocessing

The extent of missing data for each predictor is presented in Table *1*. Missingness was particularly substantial among laboratory variables, approximately 77%-79% of which were unavailable. Missing values were imputed using the median of the corresponding predictor, estimated from the training data using the SimpleImputer function in scikit-learn (NumFOCUS, Austin, TX, USA). Continuous predictors were standardised using the means and standard deviations (SDs) calculated from the training data.

Feature	Numbers (N) missing	Percentage (%) missing
Age	0	0
Sex	0	0
BMI	315	8.5
Waist circumference	435	11.8
Total cholesterol	2872	77.9
Triglycerides	2875	78
HDL-C	2876	78
Plasma glucose	2866	77.8
AST	2882	78.2
ALT	2882	78.2
GGT	3312	89.9
Platelet count	2881	78.2
Serum albumin	2882	78.2
Total bilirubin	2833	76.9
Total calcium	2833	76.9
FIB-4	2914	79.1
NFS	2917	79.2
Diabetes	0	0
Hypertension	0	0
Dyslipidemia	0	0
Female	0	0
High-risk FIB-4	0	0
High-risk NFS	0	0

TABLE 1: Extent of missing data among candidate predictors included in the cardiovascular mortality prediction models in the NHANES III MASLD cohort.

Data are presented as the number and percentage of participants with missing observations for each variable. Missingness was calculated using the total analytical cohort as the denominator (N = 3,685).

BMI: body mass index; AST: aspartate aminotransferase; ALT: alanine aminotransferase; GGT: gamma-glutamyl transferase; HDL-C: high-density lipoprotein cholesterol; NHANES III: Third National Health and Nutrition Examination Survey; MASLD: metabolic dysfunction-associated steatotic liver disease; FIB-4: Fibrosis-4; NFS: Non-alcoholic Fatty Liver Disease (NAFLD) Fibrosis Score

To minimise information leakage, imputation and standardisation procedures were fitted independently within each training fold during cross-validation and within each bootstrap resample. The corresponding preprocessing parameters were then applied to the respective validation data. Given the substantial missingness in laboratory variables and the potential limitations of single median imputation, a sensitivity analysis was performed among participants with complete laboratory data (n = 736, including 110 cardiovascular deaths).

Model development and specifications

We developed penalised Cox regression and machine-learning models to predict cardiovascular mortality. Model hyperparameters were prespecified. The model was developed using both the full set of 23 predictors and a reduced set of 18 predictors. The reduced model excluded FIB-4, the NFS, their corresponding threshold indicators, and duplicate sex coding to reduce potential collinearity.

An XGBoost time-to-event model was also developed using the survival: Cox objective. The model comprised 150 trees, a maximum tree depth of 3, a learning rate of 0.05, a subsample ratio of 0.8, a column-sampling ratio of 0.8, and a minimum child weight of 5. Censored observations were represented using negative follow-up time labels, according to the XGBoost survival:Cox implementation.

In addition, logistic regression, random forest, gradient boosting, and XGBoost classification models were developed as binary classifiers to predict cardiovascular death within 10 years (120 months). These models were evaluated only among participants whose 10-year outcome status could be determined. Participants who experienced cardiovascular death within 10 years were classified as events, whereas those who remained under observation beyond 10 years without cardiovascular death were classified as non-events. Participants who were alive but censored before 10 years were excluded because their 10-year outcome status was unknown. Participants who died from non-cardiovascular causes before 10 years were classified as non-events for the binary cardiovascular mortality outcome.

The random forest model comprised 200 trees, a maximum depth of 10, and a minimum leaf size of 10. The gradient boosting model comprised 150 estimators, a maximum depth of 3, a learning rate of 0.05, and a subsample ratio of 0.8. The XGBoost classifier was also evaluated as a 10-year binary classification model. No class re-weighting was applied.

The classification models were, therefore, distinguished from the time-to-event models in the revised analysis. In particular, the random forest and gradient boosting implementations were binary classifiers rather than survival-specific algorithms, and their descriptions have been corrected accordingly.

Prediction horizon and performance metrics

The primary prediction horizon for the binary classification models was 10 years (120 months). Model discrimination was assessed using Harrell's concordance index (C-index) for time-to-event predictions and the area under the receiver operating characteristic curve (AUROC) for 10-year binary classification. The C-index was calculated using cardiovascular death as the event of interest, with non-cardiovascular deaths censored for cause-specific analyses. The 10-year AUC was evaluated among participants with known 10-year outcome status. The 10-year Brier score was also calculated for models generating 10-year predicted probabilities.

These metrics were interpreted according to the type of prediction model and outcome being evaluated, distinguishing time-to-event discrimination from fixed-horizon binary classification performance.

Internal validation

Model performance was evaluated using three complementary approaches: apparent performance in the full development dataset, stratified five-fold cross-validation, and bootstrap optimism correction based on 200 resamples.

For five-fold cross-validation, participants were divided into five folds, stratified according to cardiovascular mortality status, using a fixed random seed of 42. Each model was fitted on four folds and evaluated on the remaining fold. Preprocessing procedures, including missing-data imputation and standardisation, were performed within the training folds to minimise information leakage.

Bootstrap optimism correction was performed using 200 resamples. For each resample, the model was fitted to the bootstrap sample, including the preprocessing procedures, and its performance was evaluated both in the bootstrap sample and in the original development dataset. The difference between these performance estimates represented the optimism for that resample. Mean optimism was subtracted from the apparent performance estimate to obtain the optimism-corrected estimate. Out-of-bag performance was also assessed. Convergence warnings and fitting failures were monitored across the full-sample, cross-validation, and bootstrap model fits.

Competing-risk analysis

Because death from causes other than cardiovascular disease precludes the subsequent occurrence of cardiovascular death, non-cardiovascular mortality was considered a competing event. Cause-specific Cox and XGBoost-Cox models were used to estimate cause-specific hazards, with non-cardiovascular deaths censored at the time of death.

To estimate the cumulative probability of cardiovascular death in the presence of competing events, we calculated Aalen-Johansen cumulative incidence functions. Participants were stratified into low-, intermediate-, and high-risk groups according to tertiles of their out-of-fold XGBoost-Cox predicted risk scores. Cumulative incidence estimates were calculated at 10 and 25 years. The corresponding 1-Kaplan-Meier estimates were also calculated for comparison, illustrating the potential overestimation of cardiovascular mortality risk when competing events are not explicitly accounted for.

Feature importance analysis

For models using tree-based methods, feature importance was estimated using gain-based metrics. We ranked predictors within each model to identify variables contributing most to discrimination.

Risk stratification

Patients were categorised into low-, intermediate-, and high-risk groups based on tertiles of model-predicted cardiovascular mortality risk. Kaplan-Meier survival curves were constructed to compare cardiovascular mortality across risk groups. Kaplan-Meier survival curves are non-parametric estimates of the probability of remaining alive (or free from a specified event) over time, accounting for right-censored observations. They provide a visual comparison of survival experiences between groups over the follow-up period.

Model performance assessment

Performance metrics included the C-index for overall discrimination and time-dependent AUC. Calibration was evaluated using calibration curves comparing predicted versus observed risks. Clinical utility was assessed using decision curve analysis across threshold probabilities from 5% to 30%.

Software

All analyses were performed using Python version 3.14.0 (Python Software Foundation, Wilmington, DE, USA). The principal packages included scikit-learn (version 1.8.0), lifelines (version 0.30.0), XGBoost (version 3.1.2), pandas (version 2.3.3), NumPy (version 2.3.5), SciPy (version 1.16.3), and Matplotlib (version 3.10.8). A fixed random seed of 42 was used to support reproducibility.

Ethical and governance considerations

De-identified data from the NHANES III, 1988-1994, linked to the National Death Index public-use Linked Mortality File was used to carry out this study [12]. Formal institutional review board approval and informed consent were not required, in accordance with institutional policy and applicable regulations

Results

A total of 3,685 adults with MASLD out of 33,994 NHANES III cohort participants met the inclusion criteria (Figure *1*). Over a median follow-up of 310 months, 504 (13.7%) cardiovascular deaths occurred, and 1,775 (48.2%) all-cause deaths occurred. The mean age (59.4 years) of participants with cardiovascular death was significantly higher than that of non-cardiovascular deaths (47.6 years), with a standard mean difference of 0.864. Participants with cardiovascular deaths had significant comorbidities like hypertension (378 (75.0%) vs 1,670 (52.5%)), diabetes (201 (39.9%) vs 765 (24.0%)), dyslipidemia (115 (2,2.8%) vs 610 (19.2%)) compared to non-cardiovascular deaths. However, patients with non-cardiovascular deaths were more overweight/obese (2,738 (86.1%) vs 407 (80. 8%)) compared to those with cardiovascular deaths.

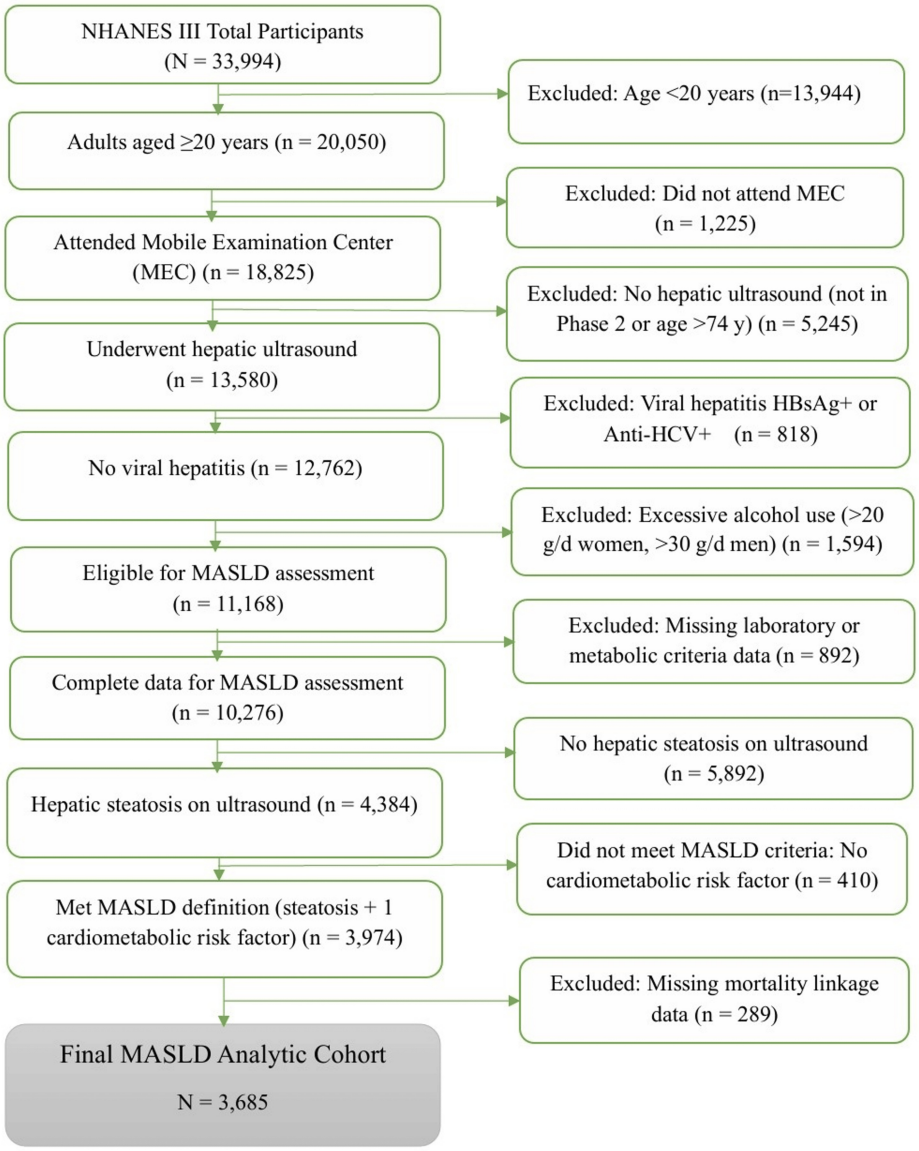


FIGURE 1: Participant selection flowchart

NHANES III: Third National Health and Nutrition Examination Survey; HbsAg: hepatitis B surface antigen; HCV: hepatitis C; MASLD: metabolic dysfunction-associated steatotic liver disease

Participants with cardiovascular mortality averaged a significantly higher total cholesterol (222.6mg/dl vs 211.9mg/dl, p=0.007), triglycerides (167.5mg/dl vs 146mg/dl, p=0.040), fasting glucose (96.5mg/dl vs 93.0mg/dl, p=0.007), and HbA1C (6.24% vs 5.74%, p<.001).

Fibrosis and risk score

MASLD patients who had cardiovascular mortality had a significantly higher risk (FIB-4 ⩾3.25) of advanced fibrosis compared to non-cardiovascular deaths (57 (11.3%) vs 150 (4.7%), SMD=0.245, P<0.001). Also, they had a high NFS (⩾0.676) compared to those with non-cardiovascular mortality (25 (5%) vs 46 (1.4%), SMD=0.201, p<0.001) (Table *2*).

Characteristic	Overall (N = 3,685)	CV Death (n = 504) (13.7%)	No Cardiovascular Death (n = 3,181) (86.3%)	SMD	P-value
Number of patients, N (%)	3685 (100)	504(13.7)	3181 (86.3)	-	-
Demographics	-	-	-	-	-
Age, years, mean ± SD	49.2 ± 15.2	59.4 ± 12.4	47.6 ± 14.9	0.864	<

					0.001*
Female sex, N (%)	1715 (46.5)	202 (40.1)	1513 (47.6)	0.151	0.002*
Race/Ethnicity, N (%)**	-	-	-	-	< 0.001*
Non-Hispanic White, N (%)	1301 (35.3)	219 (43.5)	1082 (34.0)	0.195	-
Non-Hispanic Black, N (%)	866 (23.5)	123 (24.4)	743 (23.4)	0.025	-
Mexican American, N (%)	1383 (37.5)	148 (29.4)	1235 (38.8)	0.201	-
Other, N (%)	135 (3.7)	14 (2.8)	121 (3.8)	0.058	-
BMI, kg/m², mean ± SD	30.6 ± 6.3	30.9 ± 6.2	30.6 ± 6.3	0.046	0.22
Waist circumference, cm, mean ± SD	102.5 ± 13.8	105.4 ± 14.0	102.1 ± 13.7	0.242	< 0.001*
Comorbidities and risk factors	-	-	-	-	-
Diabetes mellitus, N (%)	966 (26.2)	201 (39.9)	765 (24.0)	0.344	<0 .001*
Hypertension, N (%)	2048 (55.6)	378 (75.0)	1670 (52.5)	0.481	<0.001*
Dyslipidemia, N (%)	725 (19.7)	115 (22.8)	610 (19.2)	0.089	0.06
Overweight/Obese (BMI ≥25), N (%)	3145 (85.3)	407 (80.8)	2738 (86.1)	0.143	0.002*
Ever smoker, N (%)	1942 (52.7)	312 (61.9)	1630 (51.2)	0.216	< 0.001*
Vital signs	-	-	-	-	-
Systolic BP, mmHg, mean ± SD	129.5 ± 18.7	140.0 ± 19.4	127.9 ± 18.0	0.644	< 0.001*
Diastolic BP, mmHg, mean ± SD	77.7 ± 10.6	78.7 ± 11.8	77.6 ± 10.4	0.105	0.06
Laboratory values	-	-	-	-	-
Total cholesterol, mg/dL, mean ± SD	213.5 ± 47.6	222.6 ± 43.9	211.9 ± 48.1	0.232	0.007*
Triglycerides, mg/dL, median (IQR)	148.0 (99.0–220.0)	167.5 (102.0–238.0)	146.0 (98.0–215.8)	0.102	0.04*
HDL cholesterol, mg/dL, mean ± SD	46.9 ± 13.6	47.5 ± 13.9	46.8 ± 13.5	0.049	0.53
Fasting glucose, mg/dL, median (IQR)	94.0 (86.0–106.0)	96.5 (90.0–115.2)	93.0 (86.0–104.0)	0.180	0.007*
HbA1c, %, mean ± SD	5.81 ± 1.42	6.24 ± 1.53	5.74 ± 1.38	0.342	< 0.001*
Serum creatinine, mg/dL, mean ± SD	1.06 ± 0.26	1.11 ± 0.22	1.05 ± 0.26	0.252	0.003*
Albumin, g/dL, mean ± SD	4.26 ± 0.37	4.24 ± 0.36	4.26 ± 0.38	0.056	0.49
BUN, mg/dL, mean ± SD	14.4 ± 5.3	15.8 ± 5.3	14.2 ± 5.3	0.318	< 0.001*
ALP, U/L, median (IQR)	135.0 (123.0–152.8)	138.5 (124.5–155.8)	134.0 (123.0–152.0)	0.100	0.21
Platelet count, ×10³/μL, mean ± SD	285.5 ± 74.1	267.0 ± 71.7	288.8 ± 74.1	0.298	0.002*

C-reactive protein, mg/dL, median (IQR)	0.21 (0.21–0.55)	0.21 (0.21–0.55)	0.21 (0.21–0.49)	0.114	0.019*
Fibrosis and Risk Scores	-	-	-	-	-
FIB-4 index, median (IQR)	0.88 (0.61–1.35)	1.27 (0.83–1.80)	0.82 (0.55–1.27)	0.442	< 0.001*
NFS, mean ± SD	-1.57 ± 1.72	-0.60 ± 1.62	-1.74 ± 1.69	0.687	< 0.001*
High-risk FIB-4 (≥3.25), N (%)	207 (5.6)	57 (11.3)	150 (4.7)	0.245	< 0.001*
High-risk NFS (≥0.676), N (%)	71 (1.9)	25 (5.0)	46 (1.4)	0.201	< 0.001*
eGFR, mL/min/1.73 m², mean ± SD	69.2 ± 15.7	64.0 ± 14.8	70.1 ± 15.6	0.401	< 0.001*
Follow-up	-	-	-	-	-
Follow-up, y, median (IQR)	25.8 (16.5–28.5)	15.2 (8.6–21.8)	26.4 (19.4–28.8)	1.037	< 0.001*
All-cause mortality, N (%)	1775 (48.2)	504 (100.0)	1271 (40.0)	1.734	< 0.001*

TABLE 2: Baseline characteristics patients with MasId, stratified by cardiovascular mortality

SMD: standard mean difference; BMI: body mass index; SD: standard deviation; IQR: interquartile range; FIB-4: Fibrosis-4 Index; NFS: Non-alcoholic Fatty Liver Disease (NAFLD) Fibrosis Score; eGFR: estimated glomerular filtration rate; ALP: alkaline phosphatase; BUN: blood urea nitrogen; HDL: high-density lipoprotein. p-value <0.05 is considered significant. *: significant p-values; **Racial data for participants are presented as documented in the source dataset.

Model performance and internal validation

The performance of the prediction models differed according to the evaluation approach. Apparent performance in the full development dataset was generally higher for the tree-based models than their corresponding internally validated performance, suggesting some degree of overfitting.

In five-fold cross-validation, the C-index was 0.784 for the full penalised Cox model, 0.787 for the reduced penalised Cox model, 0.789 for XGBoost-Cox, 0.781 for logistic regression, 0.785 for random forest, 0.771 for gradient boosting, and 0.779 for the XGBoost classifier. The corresponding five-fold cross-validation AUCs for 10-year cardiovascular mortality were 0.788, 0.792, 0.796, 0.787, 0.808, 0.785, and 0.798, respectively. Although some machine-learning models demonstrated marginally higher discrimination, the differences compared with penalised Cox regression were modest (Table *3*).

Model	C-index	AUC
Cox PH (full penalized)	0.784 ± 0.007	0.788 ± 0.043
Cox PH (reduced penalized)	0.787 ± 0.008	0.792 ± 0.045
XGBoost-Cox (time-to-event)	0.789 ± 0.019	0.796 ± 0.044
Logistic regression	0.781 ± 0.013	0.787 ± 0.037
Random forest	0.785 ± 0.015	0.808 ± 0.043
Gradient boosting	0.771 ± 0.015	0.785±0.047
XGBoost classifier	0.779±0.008	0.798±0.040

TABLE 3: Five-fold cross-validation results

AUC: area under the curve; PH: proportional hazards. Values represent mean ± standard deviation across five cross-validation folds.

Bootstrap validation

The bootstrap optimism-corrected C-indices were 0.789 (95% CI: 0.771-0.805) for the full penalised Cox model, 0.792 (95% CI: 0.773-0.809) for the reduced Cox model, 0.806 (95% CI: 0.789-0.820) for XGBoost-Cox, 0.787 (95% CI: 0.770-0.803) for logistic regression, 0.804 (95% CI: 0.785-0.825) for random forest, 0.798 (95% CI: 0.780-0.816) for gradient boosting, and 0.790 (95% CI: 0.769-0.807) for the XGBoost classifier (Table *4*).

Model	Optimism-corrected C-index	C-index 95% CI low	C-index 95% CI high	AUC	AUC 95% CI low	AUC 95% CI high
Cox PH (Full penalised)	0.789	0.77	0.81	0.789	0.76	0.82
Cox PH (reduced penalised)	0.792	0.77	0.81	0.793	0.76	0.83
XGBoost-Cox (time-to-event)	0.806	0.79	0.82	0.808	0.78	0.84
Logistic regression	0.787	0.77	0.80	0.789	0.76	0.82
Random forest	0.804	0.78	0.83	0.853	0.83	0.88
Gradient boosting	0.798	0.78	0.82	0.845	0.82	0.87
XGBoost classifier	0.790	0.77	0.81	0.823	0.80	0.85

TABLE 4: Bootstrap Validation Results (95% CI)

AUC: area under the curve; CI: confidence interval; PH: proportional hazards. Bootstrap validation was performed to estimate model discrimination and uncertainty. C-index and area under the receiver operating characteristic curve (AUROC) values are reported with 95% confidence intervals derived from bootstrap resampling. AUC represents time-dependent discrimination for survival outcomes.

Overall, the internally validated results indicated broadly comparable predictive performance across the evaluated models, with no consistent evidence of a substantial advantage of the more complex machine-learning approaches over penalised Cox regression.

Feature importance

Analysis of feature importance within the XGBoost model identified hypertension and age as the most influential predictors of cardiovascular mortality, with relative importance scores of 0.15 and 0.14, respectively. Other significant contributors included female sex (0.05), diabetes (0.05), and BMI (0.04). Creatinine (0.05), total bilirubin (0.05), and fasting glucose (0.05) also demonstrated consistent predictive value. Measures of adiposity and lipid metabolism, including dyslipidemia (0.04) and waist circumference (0.04), provided additional discriminative information (Figure *2*).

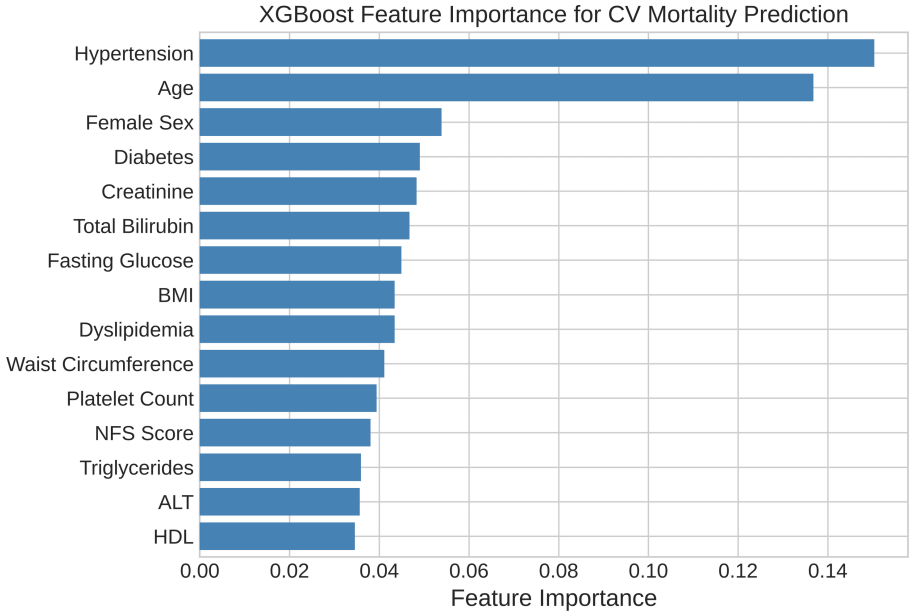


FIGURE 2: Clinical, anthropometric, and laboratory predictors for cardiovascular (CV) mortality in the XGBoost model

Relative feature importance of clinical, metabolic, and biochemical variables contributing to the XGBoost model. Hypertension and age were the strongest predictors of CV mortality, followed by female sex, diabetes, creatinine, total bilirubin, fasting glucose, body mass index (BMI), dyslipidemia, waist circumference, platelet count, NAFLD fibrosis score (NFS), triglycerides, alanine aminotransferase (ALT), and high-density lipoprotein cholesterol (HDL). Higher feature importance values indicate a greater contribution to model predictions.

Risk stratification

The model successfully stratified patients into three risk categories for cardiovascular mortality defined by tertiles. Divergence in the survival curves was observed early in the follow-up period and persisted throughout the study. Specifically, the high-risk group exhibited a sharp decline in event-free survival probability, reaching approximately 31% by 360 months. In contrast, the low-risk group maintained a near-constant survival probability exceeding 99%, while the intermediate-risk group showed a gradual decline to approximately 93% survival (Figure 3).

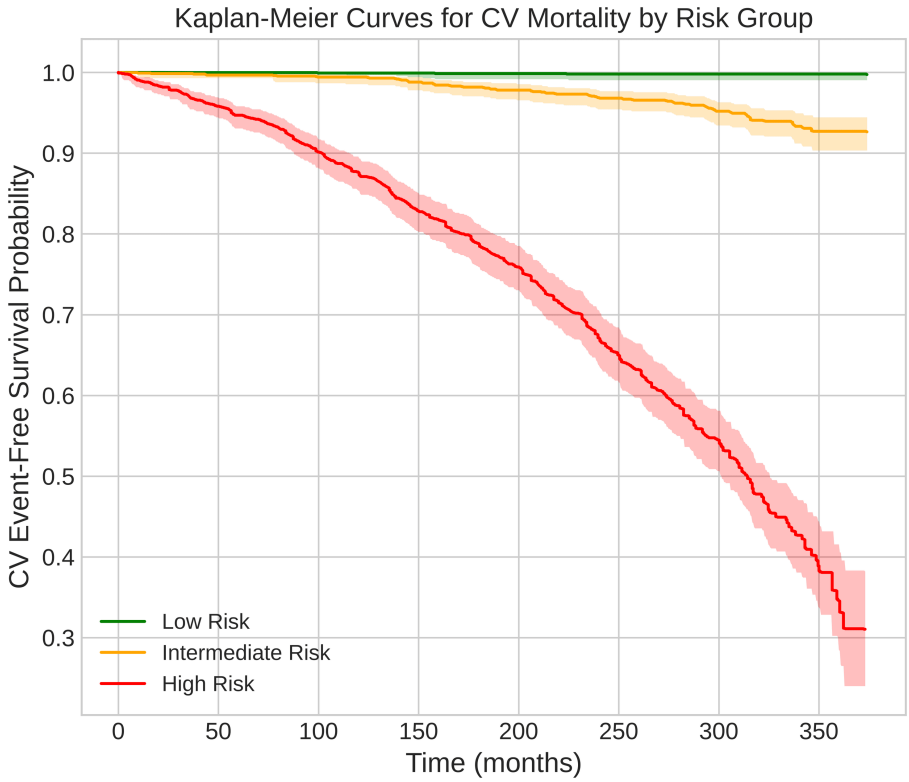


FIGURE 3: Kaplan–Meier estimates of cardiovascular (CV) event-free survival, according to risk group.

Kaplan–Meier curves for CV mortality stratified by XGBoost-derived risk groups. Kaplan–Meier estimates of CV event-free survival according to predicted low-, intermediate-, and high-risk groups. Individuals classified as high risk exhibited a markedly lower probability of CV event-free survival over follow-up compared with the intermediate- and low-risk groups. Shaded areas represent 95% confidence intervals.

Calibration and clinical utility

Calibration plots demonstrated close agreement between predicted and observed cardiovascular mortality risks for all models, with slight underestimation at the highest predicted values (Figure *4*).

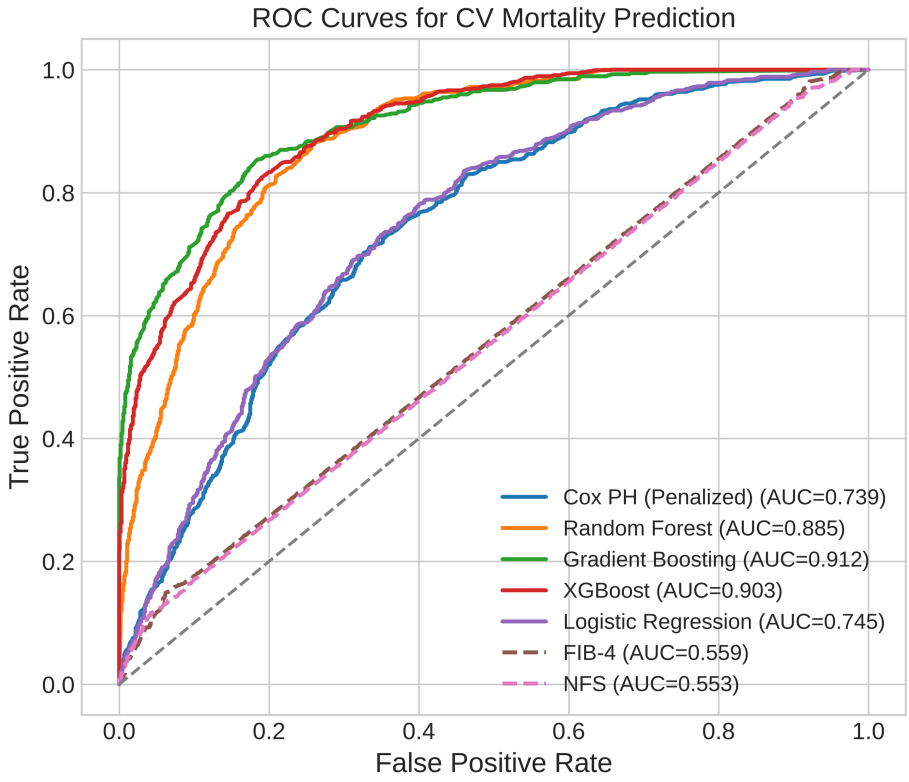


FIGURE 4: Receiver operating characteristic (ROC) curves for the prediction of cardiovascular (CV) mortality.

Comparison of the discriminative performance of machine learning algorithms and conventional risk models for cardiovascular mortality prediction. ROC curves comparing the ability of different models to predict cardiovascular mortality. Gradient boosting demonstrated the highest predictive performance (area under the curve (AUC) = 0.912), closely followed by XGBoost (AUC = 0.903) and random forest (AUC = 0.885). Conventional statistical models, including penalised Cox proportional hazards (PHs) and logistic regression, showed moderate discrimination (AUCs = 0.739 and 0.745, respectively), whereas liver fibrosis indices (Fibrosis-4 (FIB-4) Index and Non-alcoholic Fatty Liver Disease (NAFLD) Fibrosis Score) exhibited poor discriminatory performance (AUCs = 0.559 and 0.553, respectively). The diagonal reference line indicates the performance expected by random classification.

Decision curve analysis showed that machine-learning survival models offered positive net benefit across clinically relevant threshold probabilities (5%-30%) and outperformed FIB-4 and NFS throughout the threshold range (Figure *5*).

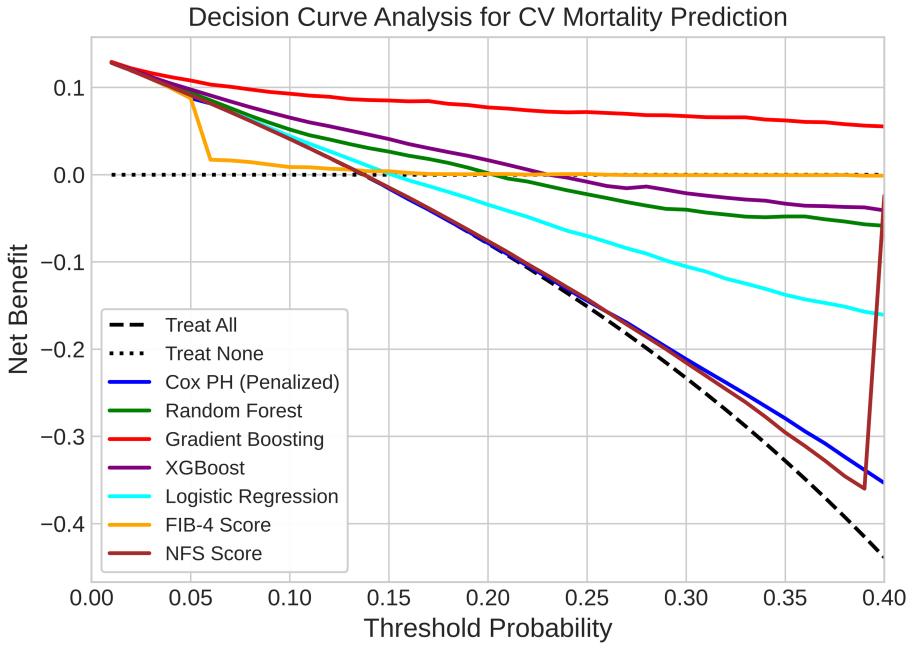


FIGURE 5: Decision curve analysis

This figure shows that machine learning models, particularly the boosting-based model, provide superior net benefit across clinically relevant threshold probabilities compared to traditional fibrosis scores, the Fibrosis-4 (FIB-4) Index and the Non-alcoholic Fatty Liver Disease (NAFLD) Fibrosis Score (NFS). This suggests improved clinical utility for cardiovascular (CV) mortality prediction.

PH: proportional hazard

Cumulative incidence of cardiovascular mortality accounting for competing risks

The cumulative incidence of cardiovascular mortality differed substantially across the three risk tertiles throughout the follow-up period. In the Kaplan-Meier analysis, which treated non-cardiovascular deaths as censored observations, the estimated cumulative probability of cardiovascular death increased progressively across the low-, intermediate- and high-risk groups, reaching approximately 5%, 17% and 52%, respectively, by 30 years. However, when non-cardiovascular death was appropriately accounted for as a competing event using the Aalen-Johansen estimator, the corresponding cumulative incidence estimates were substantially lower, at approximately 4%-5%, 13%-14% and 26%-27%, respectively. Despite the attenuation of absolute risk estimates, the risk gradient was preserved, with the high-risk tertile consistently demonstrating the greatest cumulative incidence of cardiovascular mortality, followed by the intermediate- and low-risk tertiles. These findings demonstrate the importance of accounting for competing non-cardiovascular mortality when estimating the absolute probability of cardiovascular death (Figure *6*).

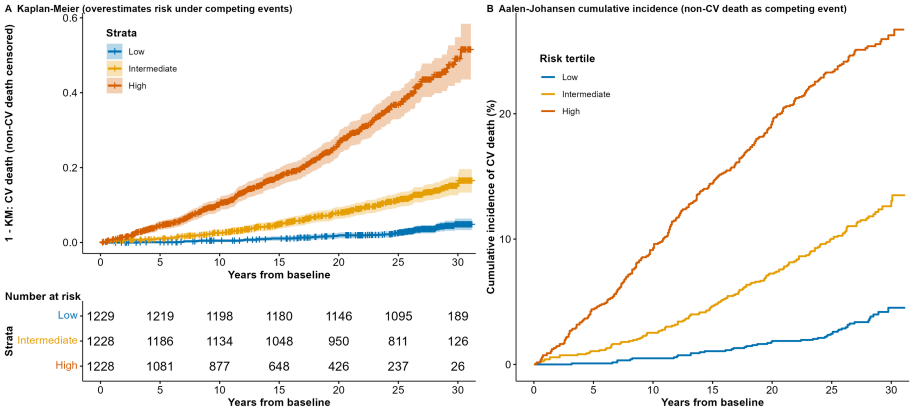


FIGURE 6: Comparison of Kaplan–Meier (KM) and Aalen–Johansen estimates of cardiovascular (CV) mortality across predicted risk tertiles.

(A) Kaplan–Meier estimates of cumulative CV mortality (1 − Kaplan–Meier), treating non-CV deaths as censored observations. Shaded areas represent 95% confidence intervals. The number of participants at risk at selected follow-up intervals is presented below the plot. (B) Aalen–Johansen cumulative incidence functions for CV mortality, with non-CV death treated as a competing event. Participants were categorised into low-, intermediate-, and high-risk tertiles based on their predicted CV mortality risk.

Discussion

Summary of findings

In this study of adults with MASLD from the NHANES III cohort, conventional statistical and machine learning survival models demonstrated varying predictive performance across internal validation approaches. Logistic regression achieved the highest discrimination during five-fold cross-validation, whereas penalised Cox regression demonstrated marginally better discrimination during bootstrap validation. However, the relatively small differences between these models do not establish a definitive performance advantage. Although gradient boosting, XGBoost, and random survival forest demonstrated high apparent discrimination when evaluated on the full training dataset, their performance was considerably attenuated during internal validation. These findings highlight the importance of distinguishing apparent model performance from validated predictive performance and suggest that the high discrimination observed with tree-based approaches may partly reflect optimism associated with evaluation on the development dataset. In a study by Chaiquan et al., it was demonstrated that machine learning algorithms significantly had better performance for reclassification to identify high-risk cardiovascular populations [16]. Machine learning tools are consistently being used to improve risk prediction, disease classification, and identification of genetic markers associated with cardiovascular diseases [17].

Interpretation and implication

In internal validation, the superior stability of logistic-regression and penalised Cox models suggests that the predictive signal is primarily driven by direct, proportional associations between risk factors and outcomes. Age and hypertension likely dominate risk because they reflect cumulative vascular injury and metabolic stress, which are central to the pathogenesis of cardiovascular mortality in MASLD [18]. Their strong and consistent contributions explain why simpler models perform nearly as well as machine-learning approaches in cross-validated settings.

Our findings suggest tree-based survival models may be leveraging nonlinear interactions and threshold effects among metabolic, hepatic, and renal biomarkers that are not easily modelled by traditional methods. Tree-based survival models have been shown to be highly effective at analysing complex biomedical data as they naturally are able to handle non-linear relationships, automatic feature interactions, and abrupt threshold effects that are often lost in traditional Cox proportional hazards models [19, 20]. Their improved clinical utility in decision-curve analysis suggests that they better identify individuals at the extremes of risk. This suggestion is supported by reports that individuals at the extremes of risk have most of their risk accounted for by combinations of abnormalities rather than from isolated predictors [21]. This highlights the multidimensional nature of cardiovascular risk in MASLD, where metabolic dysfunction, low-grade inflammation, and early organ injury converge to amplify long-term risk [21].

The clear separation of risk-stratified survival curves likely arises from the interplay of metabolic comorbidities and age-related risk accumulation, emphasising that MASLD is not merely a liver disease but a systemic cardiometabolic condition with heterogeneous trajectories [3]. Also, an important finding of this

study was the influence of competing mortality on the estimation of cardiovascular death risk. Non-cardiovascular mortality represents a competing event because individuals who die from other causes can no longer experience cardiovascular death. Our competing-risk analysis demonstrated that conventional 1-Kaplan-Meier estimates may overestimate the absolute probability of cardiovascular mortality, particularly among participants classified as having high predicted risk.

These insights have several important implications. Clinically, incorporating multidimensional risk models may enable targeted cardiovascular surveillance and early intervention for high-risk MASLD patients who may otherwise be overlooked. Methodologically, the findings underscore the importance of combining interpretable traditional modelling with more flexible machine-learning approaches to maximise predictive accuracy and clinical applicability.

Limitations and strengths

Several limitations should be considered when interpreting our findings. First, the models underwent internal validation only, and no independent external cohort was available to assess their generalisability. Although cross-validation and bootstrap optimism correction were performed, these procedures do not substitute for external validation in an independent population. The predictive performance observed in this NHANES III cohort may therefore not be directly generalisable to contemporary clinical populations.

Second, the apparent performance of some tree-based models was higher than their cross-validated and bootstrap optimism-corrected performance, suggesting potential overfitting. We therefore consider the internally validated estimates more informative than the apparent performance estimates.

Third, substantial missingness was observed among laboratory predictors, with approximately 77%-79% of values unavailable for some variables. Although median imputation was performed using training-data estimates, this approach may attenuate associations and may not fully account for the uncertainty associated with missing data. The complete-case sensitivity analysis also involved a smaller subset of participants and fewer cardiovascular deaths, which may have affected the precision of the estimates.

Fourth, non-cardiovascular mortality represents a competing event in the estimation of cardiovascular death. Although cause-specific models were used to estimate cause-specific hazards, absolute cardiovascular mortality probabilities were assessed using cumulative incidence functions. The findings demonstrate that 1–Kaplan-Meier estimates may overestimate cardiovascular mortality risk when competing events are not explicitly accounted for.

Also, the decision curve analysis suggests potential net benefit under the evaluated assumptions but does not demonstrate that implementing these models in clinical practice would improve patient outcomes.

To mitigate some of these issues, we implemented several methodological mitigations to strengthen the analysis. We used the penalised Cox proportional-hazards model and consistently reported time-dependent metrics across all models to ensure a fair comparison. To rectify the inconsistencies in model evaluation, we unified the code path for all validation schemes, ensuring that the scoring functions remained identical for both cross-validation and bootstrap procedures. By re-running the cross-validation after fixing the Cox prediction post-processing, we were able to provide a stable and transparent reporting of metrics across all models and time horizons, thereby reinforcing the validity of our comparative findings. Although externally unvalidated, rigorous internal validation using bootstrap resampling and cross-validation suggests model robustness.

Conclusions

In this nationally representative cohort of adults with MASLD, penalised Cox regression and machine learning models demonstrated moderate and broadly comparable performance in predicting cardiovascular mortality. Although some machine learning models showed marginally higher discrimination, the differences were modest and did not establish a clear predictive advantage over conventional penalised Cox regression. The competing-risk analysis further demonstrated the importance of accounting for non-cardiovascular mortality when estimating absolute cardiovascular mortality risk. These findings support the need for rigorous internal and external validation before the application of such prediction models in clinical practice. Future studies involving contemporary and independent populations are required to establish their generalisability and potential clinical utility.

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: Solomon Gyabaah, Basile Njei, Guy Nguefang, Sarpong Boateng, Yussif Issaka, Yazan

Al-Ajlouni, Gordon M. Amponsah, Ridwan A. Lawal, Ulrick Sidney Kanmounye

Acquisition, analysis, or interpretation of data: Solomon Gyabaah, Basile Njei, Sarpong Boateng

Drafting of the manuscript: Solomon Gyabaah, Basile Njei, Guy Nguefang, Sarpong Boateng, Yussif Issaka, Yazan Al-Ajlouni, Ridwan A. Lawal, Ulrick Sidney Kanmounye

Critical review of the manuscript for important intellectual content: Solomon Gyabaah, Basile Njei, Guy Nguefang, Yussif Issaka, Yazan Al-Ajlouni, Gordon M. Amponsah, Ridwan A. Lawal, Ulrick Sidney Kanmounye

Supervision: Basile Njei

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.

References

1. Kan C, Zhang K, Wang Y, et al.: Global burden and future trends of metabolic dysfunction-associated Steatotic liver disease: 1990-2021 to 2045. Ann Hepatol. 2025, 30:101898. 10.1016/j.aohep.2025.101898
2. Feng G, Targher G, Byrne CD, et al.: Global burden of metabolic dysfunction-associated steatotic liver disease, 2010 to 2021. JHEP Rep. 2025, 7:101271. 10.1016/j.jhepr.2024.101271
3. Targher G, Byrne CD, Tilg H: MASLD: a systemic metabolic disorder with cardiovascular and malignant complications. Gut. 2024, 73:691-702. 10.1136/gutjnl-2023-330595
4. Nso N, Mergen D, Ikram M, et al.: Cardiovascular morbidity and mortality in lean vs. non-lean MASLD: a comprehensive meta-analysis. Curr Probl Cardiol. 2024, 49:102569. 10.1016/j.cpcardiol.2024.102569
5. Kawanaka M, Nishino K, Kawada M, et al.: Predictors of atherosclerotic cardiovascular disease events in patients with metabolic dysfunction-associated steatotic liver disease. Intern Med. 2025, 64:1791-8. 10.2169/internalmedicine.4373-24
6. Zheng H, Sechi LA, Navarese EP, Casu G, Vidili G: Metabolic dysfunction-associated steatotic liver disease and cardiovascular risk: a comprehensive review. Cardiovasc Diabetol. 2024, 23:346. 10.1186/s12933-024-02434-5
7. Zoncapè M, Liguori A, Tsochatzis EA: Non-invasive testing and risk-stratification in patients with MASLD . Eur J Intern Med. 2024, 122:11-9. 10.1016/j.ejim.2024.01.013
8. Wang Y, Chen K, Yu J, et al.: Prognostic impact of systemic inflammation indicators on all-cause and CVD mortality in adults with MASLD. BMC Gastroenterol. 2025, 25:448. 10.1186/s12876-025-04018-3
9. Zhang F, Liu L, Li W: Developing and validating a predictive model for all-cause mortality in patients with metabolic dysfunction-associated steatotic liver disease. Diabetol Metab Syndr. 2025, 17:161. 10.1186/s13098-025-01724-6
10. Gao J, Zhang N, Aximu A, Xin N, Wang Z, Yan P: Survival machine learning models for predicting all-cause and case-specific mortality risk in metabolic dysfunction-associated fatty liver disease patients. Sci Rep. 2025, 15:42675. 10.1038/s41598-025-26729-z
11. Ngema F, Whata A, Olusanya MO, Mhlongo S: Explainable AI-driven comparative analysis of machine learning models for predicting HIV viral Nonsuppression in Ugandan patients: retrospective cross-sectional study. JMIR AI. 2026, 5:e68196. 10.2196/68196
12. National Health and Nutrition Examination Survey. (2026). Accessed: 2026: https://www.cdc.gov/nchs/nhanes/index.html.
13. Gupta U, Ruli T, Buttar D, Shoreibah M, Gray M: Metabolic dysfunction associated steatotic liver disease: Current practice, screening guidelines and management in the primary care setting. Am J Med Sci. 2024, 367:77-88. 10.1016/j.amjms.2023.11.007
14. Allen AM, Charlton M, Cusi K, et al.: Guideline-based management of metabolic dysfunction-associated steatotic liver disease in the primary care setting. Postgrad Med. 2024, 136:229-45. 10.1080/00325481.2024.2325332
15. Collins GS, Moons KG, Dhiman P, et al.: TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024, 385:e078378. 10.1136/bmj-2023-078378
16. Li C, Liu X, Shen P, et al.: Improving cardiovascular risk prediction through machine learning modelling of irregularly repeated electronic health records. Eur Heart J Digit Health. 2024, 5:30-40. 10.1093/ehjdh/ztad058
17. Singh M, Kumar A, Khanna NN, et al.: Artificial intelligence for cardiovascular disease risk assessment in personalised framework: a scoping review. EClinicalMedicine. 2024, 73:102660. 10.1016/j.eclinm.2024.102660
18. Ismail B, Nabilou P, Gluud LL, Møller S: Cardiovascular implications of MASLD: overview of the evidence .

Scand J Gastroenterol. 2025, 60:910-8. 10.1080/00365521.2025.2536585

19. Penny-Dimri JC, Bergmeir C, Reid CM, Williams-Spence J, Perry LA, Smith JA: Tree-based survival analysis improves mortality prediction in cardiac surgery. Front Cardiovasc Med. 2023, 10:1211600. 10.3389/fcvm.2023.1211600
20. Banerjee M, Reynolds E, Andersson HB, Nallamothu BK: Correction to: tree-based analysis: a practical approach to create clinical decision-making tools. Circ Cardiovasc Qual Outcomes. 2019, 12:004879-10. 10.1161/HCQ.0000000000000056
21. Michalopoulou E, Thymis J, Lampsas S, et al.: The triad of risk: linking MASLD, cardiovascular disease and type 2 diabetes; from pathophysiology to treatment. J Clin Med. 2025, 14:428. 10.3390/jcm14020428