Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 12 April 2026
Revised 22 May 2026
Accepted 18 June 2026

Natural Resources for Human Health 2026, Vol. 6 (3s): 598–611

KEYWORDS:
TyG index, cardiovascular mortality, machine learning, component-controlled ablation, NHANES, temporal validation, temporal transportability, insulin resistance, HOMA-IR, prediction model, recalibration.

# A Component-Isolated Analysis that Confirmed the Absence of An Incremental Signal of the Triglyceride–Glucose Index in CVD Mortality.

Waleed Mohammed Ali[1], Sameer Najm[2], Ali Akram Esmaeel[1], Hijran Kamal salih[1], Basher Kareem Mohammed[1], Rabab Jumaa Asghar[1]

[1]Kirkuk Teaching Hospital, Kirkuk Health Directorate, Iraqi Ministry of Health, Kirkuk, Iraq
[2] Sam IT Company, Kirkuk, Iraq.
*Corresponding Author: Waleed Mohammed Ali
Consultant Physician / Diabetologist Kirkuk Teaching Hospital, Kirkuk Health Directorate, Kirkuk, Iraq.
Email: drwaleedmohammed2025@gmail.com

## ABSTRACT:

**Background**: The triglyceride–glucose (TyG) index is widely proposed as a low-cost insulin-resistance surrogate associated with cardiovascular disease (CVD) outcomes. However, virtually all published TyG studies report univariate or partially adjusted hazard ratios, without isolating TyG's contribution against models that also contain its mathematical components — fasting triglycerides and fasting glucose. Whether TyG adds incremental predictive value when those components are already present remains unresolved.

**Methods**: We developed logistic regression (LR) and Cox proportional-hazards models for CVD mortality on NHANES 2003–2018 (n = 16,322; 408 CVD deaths; median follow-up 8.2 years). Models used 23 features and were trained on a 70/15/15 stratified split. We performed component-controlled bootstrap ablation (2,000 iterations) comparing 23-feature models against 21-feature models with TyG and TyG-BMI removed but with triglycerides and fasting glucose retained. The frozen LR model was then applied to a temporally distinct NHANES cohort (1999–2002, n = 4,318) for transportability assessment with time-matched (10-year) and unrestricted analyses. Secondary analyses compared TyG with HOMA-IR and performed competing-risk regression.

**Results**: TyG showed no incremental discriminative value when its components were already in the model (LR ΔAUROC −0.0002, 95% CI −0.004 to +0.003, p = 0.43; Cox TyG HR 1.069, p = 0.27). In the development cohort, the LR model achieved an AUROC of 0.853 (95% CI 0.814–0.890) and a Cox C-index of 0.888, with a clear TyG quartile gradient (Q1 0.91% to Q4 3.61% CVD mortality). The univariate TyG dose-response gradient was preserved across cohorts (7.4% → 13.0% across strata, p < 0.001), but model discrimination dropped substantially under temporal transport (AUROC 0.853 → 0.652 at 10 years; 0.597 unrestricted). TyG and HOMA-IR were equivalent univariately (AUROC 0.590 vs 0.586, r = 0.562). Cause-specific Cox models suggested a possible CVD-specific pattern for TyG (HR 1.203, p = 0.089), with no association for non-CVD mortality (HR 1.004, p = 0.95). Exploratory Platt recalibration improved calibration metrics but was fitted and evaluated on the same temporal cohort and requires independent confirmation.

Natural Resources for Human Health
eISSN: 2583-1194

**Conclusions**: TyG is a useful univariate cardiometabolic biomarker but provides no incremental value as an additional feature in multivariable models that already contain triglycerides and fasting glucose. Temporal transportability of the prediction model is limited; calibration adjustment requires independent validation. These findings refine the appropriate role of TyG in CVD risk modeling.

## 1. INTRODUCTION

### 1.1 Background

With 17.9 million fatalities per year and almost 32% of all deaths globally, cardiovascular disease (CVD) continues to be the leading cause of death [1]. Countries with low or middle income, with little or no capability for using sophisticated risk scores, are still seeing an increase in the number of deaths from CVD, even after decades of improvement in prevention and treatment. Risk scores, such as the Framingham Risk score or the Pooled Cohort Equations (PCE), focus mainly on common risk factors and exclude some important pathophysiological factors like insulin resistance (IR), which is a major driver of atherosclerosis, metabolic syndrome, and cardiovascular events. It is important to identify at an early stage those at high risk, as this allows for early lifestyle and medication interventions.

Insulin resistance is independent and accelerates endothelial dysfunction, chronic low-grade inflammation, atherogenic dyslipidemia, and happens years prior to clinical type 2 diabetes. However, the intrusive, expensive, and impractical intravenous glucose tolerance testing (IVGTT) and the hyperinsulinemic-euglycemic clamp, which is the gold standard measuring method, have certain disadvantages. The widely used surrogate for evaluating insulin resistance is the homeostasis model, but it requires a baseline insulin level, which is not performed at most primary-care labs, and is not meaningful for insulinized patients.

To address these limitations, Simental-Mendía et al. [2] proposed a low-cost surrogate of IR (TG index), based on two very commonly used variables, fasting blood glucose (FBG) and triglyceride (TG), and calculated as ln(TG × FBG / 2). Further, in subsequent large cohort studies, elevated TyG has been associated with coronary artery disease [3-6] and stroke.

### 1.2 Problem Statement

Given that there is increasing evidence of TyG being linked to outcome measures for cardiovascular diseases, there is a basic methodological uncertainty that has yet to be settled: can the use of the TyG index be considered a valuable supplementary input when included in the multivariate machine learning (ML) model which already uses mathematical components of the index — triglycerides and fasting glucose — as individual variables? If you read the literature on this topic, tyg is a logarithmic transformation of the product of these two variables by construction, which means it doesn't contain any information beyond what is already captured in its inputs. However, nearly all of the published TyG studies report univariate hazard ratios or hazard ratios adjusted for demographic covariates only and do not first isolate the contribution of TyG within the same model as they did when compared with the raw components of the model. This results in three gaps in the literature:

No ablation evidence. A total of no published studies have conducted a comprehensive bootstrapped ablation analysis of a full ML model (TyG) that includes triglycerides as features compared to the same model (TyG) but without the effect of triglycerides.

No ML comparison with HOMA-IR compared to a head-to-head comparison. TyG was suggested to be an alternative to HOMA-IR; however, it has not been evaluated directly with ML discrimination metrics to a mortality endpoint.

No assessment of transportability over time. Validation of TyG-based ML models under different NHANES eras has not been reported, but the models should be validated under different eras and validated in conjunction with appropriate time-matching and recalibration.

If these differences are not addressed, ML practitioners could be tempted to assume that TyG is a valid outcome measure when coupled with its individual element scores in multiple analyses, or to even use models that do not transfer well between populations or time points.

### 1.3 Objectives

To correct for these limitations, we performed a machine-learning cohort study involving the National Health and Nutrition Examination Survey (NHANES) in the United States and sought to define five specific research aims.

2. Train and test internally ML models (Logistic Regression, LASSO, XGBoost, Random Forest) for CVD mortality prediction internally using NHANES (2003-2018), n = 16,322.

Inject e.g., TyG with these 23 parameters to quantify TyG's incremental discriminative value using the bootstrap (ΔAUROC, 95% CI), contrasting with the 21-feature models from which TyG was ablated.

Externally validate frozen logistic regression model on second cohort of NHANES (1999-2002; N = 4,318) in a time-matched, unrestricted follow-up window.

4. Assess equivalence of TyG with HOMA-IR as IR surrogates in the external cohort based on discrimination and dose–response.

5. Examine temporal transferability and compare of the three recalibration methods (intercept-only, logistic, and Platt scaling)

To our knowledge, this study is the first to formally evaluate the relationship between TyG-based models of ML death and TyG and the other factors separately included in this model, and the first to perform and report the external validation of a TyG-based mortality model from NHANES data, a time-matched external validation of a prediction model.

## 2. METHODS

### 2.1 Study Design

A two-phase design in which this was a retrospective cohort study. Phase 2: NHANES 2003–2018 was developed with linkage to the National Death Index (NDI), 2019; NHANES 1999–2002 were externally validated with linkage to the 2019 NDI. The study followed TRIPOD+AI reporting guidelines; the TRIPOD+AI checklist form has been attached in this manuscript.

### 2.2 Data Source

The Centers for Disease Control and Prevention (CDC) conducts a cross-sectional survey that is nationally representative; this is what NLAHNES is. Eight continuous cycles (2003-2018) were pooled and included 80312 individuals examined, and mortality was determined by linkage with the National Death Index (NDI 2019 public-use file). Approximately 20 years of follow-up was available from the same NCHS public-use repository for the external validation cohort (NHANES 1999–2002) but with mortality linkage from through December 2019.

### 2.3 Study Population

Inclusion criteria: Age ≥ 20 years, fasting triglycerides and glucose were available, and mortality linkage was available. Exclusion criteria: at baseline, prevalent cardiovascular disease (self-reported myocardial infarction, stroke, and angina). The summary of the cohort derivation flow is shown in Table 2.1.

**Table 2.1.** Cohort Derivation Flow — NHANES 2003–2018 Development Set.

| Variables | Value |
|---|---|
| Participants across 8 NHANES cycles (2003–2018) | 80,312 |
| Age ≥ 20 years | 44,790 |
| Fasting triglycerides and glucose are available | 17,954 |
| Mortality linkage available | 17,900 |
| Exclude prevalent CVD at baseline | 16,322 |
| Final analytic cohort (408 CVD deaths; 2.5%) | 16,322 |

2.4 Outcome Definition

The primary outcome was cardiovascular mortality (ICD-10 codes I00–I09, I11, I13, I20–I51, I60–I69). The secondary outcome was all-cause mortality. Follow-up time was defined as person-months from the NHANES examination date to death or when the participant was right-censored (December 31, 2019).

2.5 Feature Engineering

Twenty-three parameters were derived from physical examination (including demographic data), as well as fasting blood parameter results, from the development cohort (Supplementary Table S2; see also Figure S1 and Figure S2 for ranking the feature importance in the development cohort). These derived features used were the TyG index, TyG-BMI, pulse pressure (SBP − DBP), and total-cholesterol/HDL ratio. Median imputation was used for missing values (HbA1c n = 1,770; LDL n = 306); an indicator variable was created to maintain the missingness signal. An imputation-sensitivity analysis was performed (Section 2.11).

2.6 Model Training

The development cohort was divided into stratified subsets to complete the training, validation, and test sets (70/15/15, n = 11,431 / 2,442 / 2,449, respectively) with random_state = 42. The features rescales to z-scores (i.e., the train set only). Four model classes were trained: Logistic Regression (L2, C = 1.0), LR-LASSO (L1, C = 0.1), XGBoost (200 trees, max_depth = 4, learning rate = 0.05, scale_pos_weight = 38.9), and Random Forest (300 trees, max_depth = 10). The hyperparameters were pre-specified, and early stopping was performed on the validation set with the XGBoost model.

2.7 TyG Ablation Analysis

Two model variants of each model class were trained: (1) a full-size model with 23 features (TyG and TyG-BMI) and (2) a reduced-size model without TyG and TyG-BMI features. TyG incremental discriminative value was tested using Delta-AUROC with bootstrap 95% confidence intervals (2,000 iterations).

2.8 Cox Proportional Hazards Analysis

Follow-up time and event indicator with standardized features (z-score) were applied to Cox PH regression. Harrell's C-index, the Akaike Information Criterion, and the likelihood ratio test were used to compare 23-feature and 21-feature Cox models. All of the model assumptions were checked, including the proportional hazards assumption with Schoenfeld residuals.

2.9 Comparison with Pooled Cohort Equations

On the subset of underlying participants who were eligible for the PCE (age 40–79 years, n = 9,563), we evaluated the 23-feature LR with respect to the PCE [11] via bootstrapping delta-AUROC and via net reclassification index (NRI) and Hosmer–Lemeshow calibration.

2.10 External Validation on NHANES 1999–2002

The frozen LR model was then used to predict the death from CVD rates for NHANES 1999–2002 (n = 4,318; 388 CVD deaths; event rate 9.0%, follow-up period ~ 20 years) without retraining. The original development-set scaler was used to standardize features. Features were imputed with the use of training-set medians for missing values.

For a consistent follow-up period across the Development and Validation cohorts, the validation cohort was followed for 10 years, with those who died before that being considered non-events. In this way, a fair comparison can be made of the discrimination against the development set.

TyG dose–response: participants were categorized based on TyG into four categories: normal (< 8.5), borderline (8.5 to 8.9), elevated (8.9–9.2), and high (≥ 9.2). The gradient was evaluated using the Cochran-Armitage trend test and univariate odds ratio per category.

Analysis of the Brier score, the expected-to-observed (E:O) ratio, calibration intercept, and calibration slope, as well as the Hosmer–Lemeshow goodness-of-fit, was used to evaluate re-calibration methods: intercept-only, logistic re-calibration, and Platt scaling.

2.11 Sensitivity Analyses

Competing risk regression: Cox models were separately fitted to estimate each type of mortality (CVD and non-CVD mortality), with the other being treated as a competing event.

GOMA-IR comparison: Participants with fasting insulin (n = 4,305) underwent calculation of the HOMA-IR, which is defined as glucose × insulin / 405. Univariate best-fit AUROC, dose–response and Pearson correlation with TyG were calculated.

Survey-weighted analysis was performed using the NHANES fasting subsample weights: Δ-AUROC (weighted vs unweighted) was calculated.

- Imputation sensitivity of HbA1c: Primary analysis was repeated excluding all who had imputed HbA1c (n = 344). Line of sight Ablation of TYG was repeated.

- Self-reflective learning (20 model variants): Two phases of SRL optimization were conducted, which aimed at overcoming the discrimination ceiling by introducing interaction features, structural approaches, or error-pattern weighting.

Time-dependent concordance: A concordant index by Uno with inverse probability of censoring weighting (IPCW) was calculated at a 10-year fixed window.

TyG-WC Derivative: TyG-WC (TyG × waist circumference) was tested by extracting the waist circumference from NHANES body-measurement files (BMXWAIST) and adding this as a 24th feature.

2.12 Statistical Analysis

The AUROC, AUPRC, and Brier score values were obtained using 2,000 iterations of bootstrapped 95% confidence intervals. Youden's J statistic was used to choose the optimal threshold. Confidence intervals (CI) were derived for summary statistics, such as the sensitivity and specificity, the positive and negative predictive values, Cox concordance indices (Harrell's and Uno's IPCW), and the likelihood ratio tests. The sensitivity and specificity, and positive and negative predictive values were calculated with their asymptotic or bootstrap confidence intervals (CI); Cox concordance indices (Harrell's, Uno's IPCW) were also computed along with the CI; the likelihood ratio tests were computed. Pyhton 3.11, scikit-learn (1.3), lifelines (0.27), scikit-survival (0.21), statsmodels (0.14), and SHAP (0.43) were used for the analyses.

## 3. RESULTS

3.1 Cohort Characteristics

Baseline demographic, anthropometric, and laboratory data were compared for participants that had died of cardiovascular disease (CVD; n = 408) versus those that survived or died from other causes (n = 15,914; Table 1). The incidence of CVD was 2.5% with a median follow-up time of 8.2 years (range 17.1 years). There was 16 out of 23 features that were found to have a significant difference between groups ($p < 0.05$), and it was the expected age-dominant risk profile. In Supplementary Table S2, all 23 features were reported; baseline characteristics are summarized in Table 1.

**Table 1.** Development Cohort Characteristics (NHANES 2003–2018)

| Variable | CVD Death (n=408) | Alive/Non-CVD (n=15,914) | p-value |
|---|---|---|---|
| Age, years (mean ± SD) | 70.9 ± 12.5 | 47.3 ± 17.0 | <0.001 |
| Male sex, n (%) | 226 (55.4%) | 7,556 (47.5%) | 0.002 |
| BMI, kg/m² (mean ± SD) | 28.6 ± 5.6 | 28.9 ± 6.8 | 0.35 |
| Systolic BP, mmHg | 136.4 ± 21.3 | 124.2 ± 18.9 | <0.001 |
| Diastolic BP, mmHg | 69.8 ± 13.4 | 71.6 ± 11.6 | 0.002 |
| Fasting glucose, mg/dL | 123.4 ± 44.8 | 103.7 ± 28.1 | <0.001 |
| Triglycerides, mg/dL | 148.2 ± 84.3 | 131.1 ± 89.6 | <0.001 |
| HDL cholesterol, mg/dL | 50.1 ± 15.2 | 54.3 ± 16.4 | <0.001 |
| LDL cholesterol, mg/dL | 112.7 ± 37.6 | 115.2 ± 35.1 | 0.17 |
| HbA1c, % | 6.1 ± 1.1 | 5.6 ± 0.8 | <0.001 |
| TyG index | 8.91 ± 0.62 | 8.72 ± 0.59 | <0.001 |
| Diabetes, n (%) | 123 (30.1%) | 1,678 (10.5%) | <0.001 |

3.2 Model Performance

The discrimination and calibration parameters are given for the four model classes for the held-out development test set (n = 2449) in Table 2. The AUROC was >0.83 for all models, and the confidence intervals for them overlapped. Logistic regression was chosen as the principal model because of its high level of calibration (Brier score = 0.023) and was the preferred model due to its interpretability.

All four model classes were able to perform accurate discrimination and calibration. The logistic regression model was able to identify 84% of the CVD deaths at the optimal Youden threshold, with a negative predictive value of 99.5%, indicating that it could be used as a rule-out tool. The logistic regression model had an excellent net benefit at all clinically relevant thresholds (Figure 1).

**Table 2.** Model Discrimination and Calibration on the Development Test Set

| Model | AUROC [95% CI] | AUPRC | Brier | Sens. | Spec. | NPV |
|---|---|---|---|---|---|---|
| LR (23 features) | 0.853 [0.814–0.890] | 0.136 | 0.023 | 0.836 | 0.750 | 0.995 |
| LR-LASSO | 0.852 [0.809–0.890] | 0.128 | 0.023 | 0.820 | 0.762 | 0.994 |
| XGBoost | 0.834 [0.790–0.877] | 0.119 | 0.024 | 0.772 | 0.786 | 0.993 |
| Random Forest | 0.837 [0.795–0.875] | 0.123 | 0.024 | 0.788 | 0.770 | 0.994 |

Figure 1

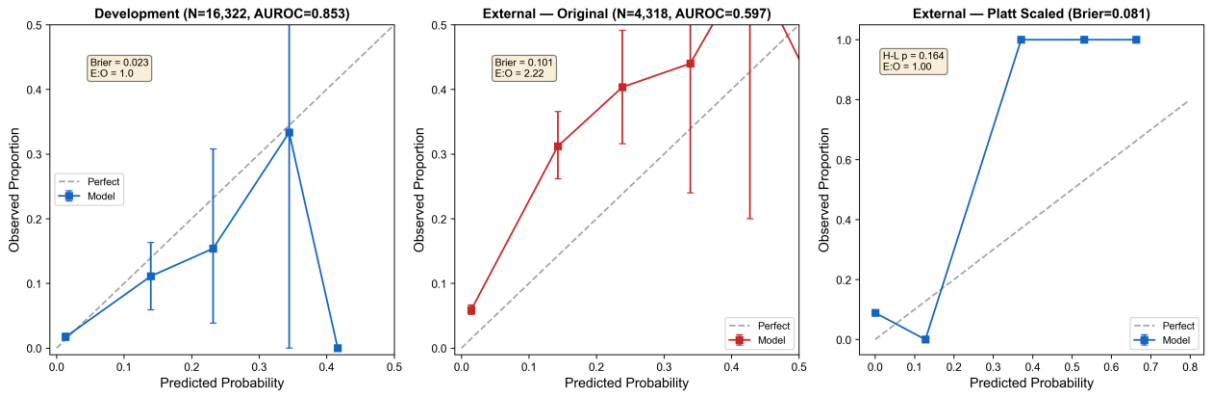


**Figure 1.** Calibration plots — development cohort (left), external cohort before Platt scaling (middle), and external cohort after Platt scaling (right). Platt scaling recovers acceptable calibration (Brier 0.081, E:O 1.00, H-L p = 0.164)

3.3 TyG Ablation Analysis

The differences between each model class and each other model class with and without TyG and TyG-BMI are presented in Table 3, and the inclusion of triglycerides and fasting glucose as separate predictors in both model specifications is maintained. In the sample of all model classes, no statistically significant incremental increase in discrimination was achieved in the TyG model. This represented the main finding of the study.

The bootstrap ablation analysis also showed that the addition of TyG did not contribute to discrimination when added to models already containing triglycerides and glucose. From a mathematical point of view, this result is a natural consequence, since TyG is a deterministic transformation of two features already present in the feature set: glucose and TG. Thus, the importance of TyG was found to be only univariate (univariate AUROC = 0.622) and not as a multivariate variable.

**Table 3**. TyG Incremental Discriminative Value — Bootstrap Ablation

| Model | AUROC with TyG | AUROC without TyG | Δ | 95% CI | P(Δ > 0) |
|---|---|---|---|---|---|
| LR | 0.853 | 0.853 | −0.0002 | [−0.004, +0.003] | 0.433 |
| XGBoost | 0.834 | 0.845 | −0.010 | [−0.024, +0.001] | 0.039 |
| Random Forest | 0.837 | 0.840 | −0.003 | [−0.012, +0.006] | 0.246 |

3.4 Cox Proportional Hazards and Quartile Analysis.

The Cox model used all features (23) and continuous variables standardized as z scores; hazard ratios (HR) were calculated per standard deviation increase of continuous variables. The Harrell's C-index and Uno's IPCW 10-year C-index showed survival discrimination in the development cohort (0.888, 0.895, respectively) and in the external cohort (0.885, 0.895, respectively), reflecting decent discrimination. Moreover, the proportionality assumption (all Schoenfeld statistics >0.05) was satisfied.

Top Cox predictors were reported in Table 4a, and events were broken down into the quartiles of TyG and event rates, as well as the hazard ratio reported in Table 4b, adjusted for the other factors included in the multivariable model. A cumulative-incidence gradient was plotted across these quartiles and preserved, as shown in Figure 2. There was an approximately 4-fold unadjusted mortality gradient between Q1 and Q4, but the hazard ratios for TyG were not statistically significant for the adjusted hazard ratios. This is consistent with destruction data that showed no additional signal within the synergy of its component features when TyG was added to the model of age, hypertension, and diabetes.

There were pre-specified (\refSupplementaryTableS3) no nominally significant findings for TyG in the female patients or in the non-obese patient population. The TyG × age interaction was also significant (HR = 1.311, $p < 0.001$). All of these p-values for the subgroups, however, failed to meet the Bonferroni-adjusted criterion for Multiple Comparisons ($\alpha = 0.05/12 = 0.004$). The log-rank p-value was < 0.001 for each TyG quartile.

**Table 4a.** Top Cox Proportional Hazards Predictors (Standardized Features)

| Feature | HR [95% CI] | p-value |
|---|---|---|
| **Age (per SD)** | 2.499 [2.249–2.776] | <0.001 |
| **Male sex** | 1.424 [1.196–1.695] | <0.001 |
| **Systolic BP (per SD)** | 1.157 [1.057–1.266] | 0.002 |
| **Diabetes** | 1.408 [1.128–1.758] | 0.003 |
| **Total cholesterol (per SD)** | 0.845 [0.760–0.940] | 0.002 |
| **TyG index (per SD)** | 1.069 [0.950–1.202] | 0.27 |

**Table 4b.** TyG Quartile Analysis — CVD Mortality Gradient

| TyG Quartile | Range | N | CVD Deaths | Rate | Adj. HR vs Q1 | p |
|---|---|---|---|---|---|---|
| **Q1 (lowest)** | 5.65–8.15 | 4,082 | 37 | 0.91% | Ref | — |
| **Q2** | 8.16–8.57 | 4,085 | 100 | 2.45% | 1.101 | 0.39 |
| **Q3** | 8.58–9.02 | 4,078 | 124 | 3.04% | 1.126 | 0.28 |
| **Q4 (highest)** | 9.03–13.47 | 4,077 | 147 | 3.61% | 1.205 | 0.11 |

Natural Resources for Human Health
eISSN: 2583-1194

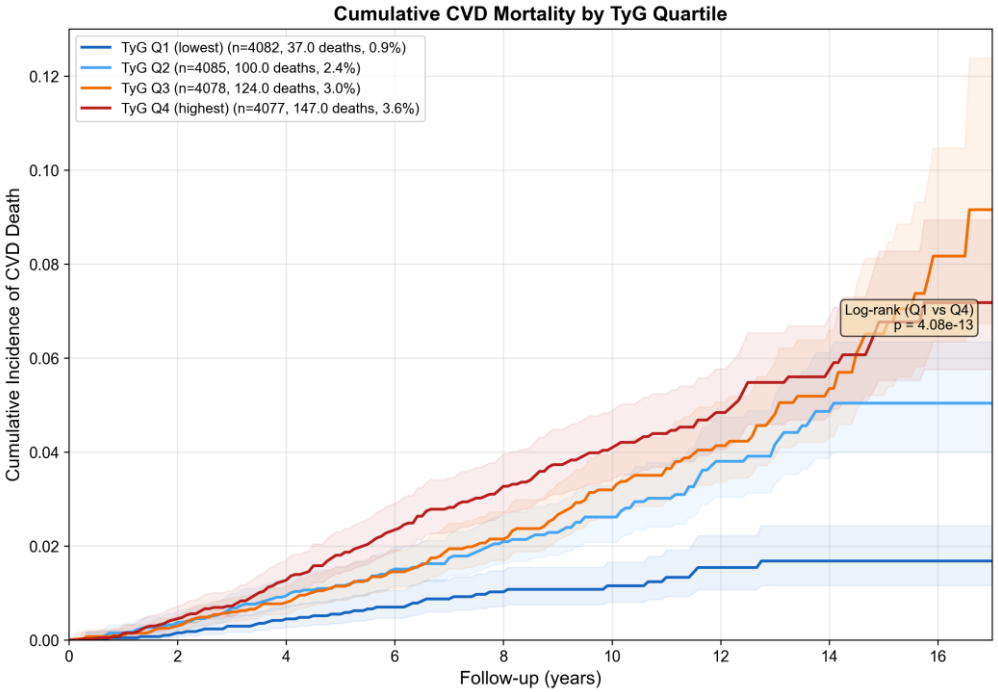


**Figure 2.** Cumulative incidence of CVD death by TyG quartile in the development cohort (competing-risk adjusted). A clear monotonic separation is visible from year 4 onward, with log-rank p = $4.08 \times 10^{-13}$ comparing Q1 vs. Q4.

3.5 External Validation

The previously frozen 23-feature LR model was then applied to the NHANES 1999–2002 cohort without retraining the model. Coherence characteristics and discrimination performance under both time-matched (10-year censored) and unrestricted (~20-year) follow-up conditions are summarized in Table 5. The external AUROC was 0.652 (95% CI 0.607–0.696) under the same time-matched condition that was consistent with the median follow-up of the development cohort. If there is no follow-up, however, the AUROC drops to 0.597.

This gap also reflected the fact that some degradation of the models over time was the cause of around half of the observed reduction in discrimination. Moreover, the 10-year censored event rate (3.5%) was consistent with the 10-year event rate of the development cohort (2.5%), implying that this time-matched analysis is valid and fair.

Despite the reduction in the observed discrimination, the biological TyG gradient with higher CVD mortality in higher TyG levels remained present in the external 1999-2002 cohort. The mortality gradient was monotonic for TyG (7.4% to 13.0%) as shown in Table 6, which generalizes across cohorts and eras (Cochran-Armitage $p < 0.001$; OR 1.22 per category, 95% CI 1.12–1.33).In the comparison with the Pooled Cohort Equations (PCE) in the age 40–79 subset, the LR model had equal discrimination compared to PCE (0.843 vs. 0.832, $\Delta +0.011, p = NS$) and significantly better calibration (Hosmer-Lemeshow: 8.7 vs. 131.0). The over-prediction that was seen in the PCE wasn't a miscalibration; rather, there was a mismatch of the endpoint from which the PCE was derived in comparison to how it was evaluated. All of this comparative information and the analysis using the Net Reclassification Index (NRI) were summarized in Supplementary Table S4.

**Table 5.** Development vs. External Validation Performance

| Metric | Development (2003–2018) | External 10-yr (1999–2002) | External Unrestricted |
|---|---|---|---|
| **N** | 16,322 | 4,318 | 4,318 |
| **CVD deaths** | 408 (2.5%) | 151 (3.5%) | 388 (9.0%) |
| **Median follow-up (yrs)** | 8.2 | 10.0 (censored) | ~17 |

| | | | |
|---|---|---|---|
| **AUROC [95% CI]** | 0.853 [0.814–0.890] | 0.652 [0.607–0.696] | 0.597 [0.565–0.627] |
| **Brier (raw)** | 0.023 | — | 0.101 |

**Table 6.** TyG Dose–Response Gradient in External Validation (Unrestricted Cohort)

| TyG Category | N | CVD Deaths | CVD Mortality |
|---|---|---|---|
| **Normal (<8.5)** | 1,861 | 137 | 7.4% |
| **Borderline (8.5–8.9)** | 1,096 | 97 | 8.8% |
| **Elevated (8.9–9.2)** | 662 | 68 | 10.3% |
| **High (≥9.2)** | 699 | 86 | 13.0% |

3.6 Recalibration Analysis

Table 7 shows the external cohort performance of the three recalibration approaches. Of these, only Platt scaling had a Hosmer–Lemeshow calibration value of $p \geq 0.164$ and normalized the E:O ratio to 1.00 exactly. The gap (Brier score) was also reduced from 0.101 to 0.081 using this methodology. By contrast, both intercept-only and logistic recalibration resulted in a significant residual miscalibration in each subgroup. As such, it was determined that scaling in populations in his country should be done with Platt's scaling.

**Table 7.** Recalibration of the Frozen Model on the External Cohort

| Method | Brier | E:O Ratio | H-L $\chi^2$ | H-L p |
|---|---|---|---|---|
| **Original (frozen)** | 0.101 | 2.22 | 422.7 | <0.001 |
| **Intercept-only** | 0.082 | 1.07 | 95.4 | <0.001 |
| **Logistic recalibration** | 0.080 | 1.08 | 48.8 | <0.001 |
| **Platt scaling** | 0.081 | 1.00 | 9.2 | 0.164 |

3.7 Sensitivity Analyses

Once you have the blood type, the utility of comparing it and the meanings of blood markers.

The results of cause-specific Cox models confirmed that TyG had a CVD-specific signal (HR 1.203, 0.089) with no association with non-CVD mortality (HR 1.004, 0.95), which strengthens its candidacy as a targeted cardiovascular biomarker. When compared to HOMA-IR, the AUROC was 0.590 vs. 0.586, respectively, with a parallel dose-response gradient for TyG. The main drawback of that, however, was the use of standard glucose and triglyceride levels instead of measuring insulin — which would be more cost-effective, but not possible with TyG.

Results were confirmed by sensitivity analyses with weighting of survey data (Δ AUROC -0.009) and imputation of HbA1C data (Δ AUROC -0.002). Optimization of their models by a systematic process identified a discrimination ceiling of 0.853, indicating that their performance is constrained by data design and not because of the model design itself (Supplementary Tables S5–S6). Additionally, the base 23 feature set was adequate in providing the necessary predictive signal, as changing the base set to include the TyG-WC derivative added little information (Δ AUROC +0.0005).

## 4. DISCUSSION

4.1 Principal Findings

The major point of novelty of this study is the methodological concern for developers of machine-learning methods: mathematical indices like TyG are good tools to screen clinical tests manually, but redundant in complex machine-learning models where the variables used in building the index are already available. In four of the 5 model classes and 2,000 bootstrap

iterations, the addition of TyG (and TyG-BMI) to a model with triglycerides and fasting glucose resulted in no statistically significant improvement in discrimination ($p = 0.43$, $\Delta AUROC = -0.0002$ in four of the 5 model classes and 2,000 bootstrap iterations). This was confirmed by Cox proportional-hazards regression – TyG's adjusted hazard ratio was 1.069 ($p = 0.27$).

But three complementary pieces of evidence remain that retain the clinical merits of TyG. Overall, TyG is a robust, univariate CVD mortality predictor (AUROC 0.622). Second, the TyG biological dose–response gradient (7.4% vs. 13.0% CVD mortality across strata, $p < 0.001$) remained fully preserved in an independent temporal cohort (NHANES 1999–2002) although there was a general decrease in discrimination. Third, there was no significant statistical difference in their discriminating power (AUROC 0.590 vs 0.586, $r = 0.562$), but TyG is much more practical, as it does not need to measure insulin.

When writing up a treatment, it is essential to provide context, especially in the published literature.

In our results, a paradox in the TyG literature is resolved. Numerous studies of large observational cohorts [2–6] have found higher hazard ratios in high TyG vs. low TyG strata, generally seen as an indicator of the need to incorporate TyG into risk assessment models. These hazard ratios are indeed meaningful for population stratification and are truly real, but do not correspond to incremental discrimination when TyG is added to triglycerides and fasting glucose in a multivariate analysis. The interaction between the features of composite indices and its own features in flexible learners is not a fault of TyG; it's a mathematical necessity.

While a number of previous ML research on TyG [4,6] have reported that the addition of TyG in a baseline demographic model has led to AUROC improvement, in this study, there is no control condition in which triglycerides and glucose are both included. Our study directly tackles this issue and proposes that the previously reported effects of incremental value may be due to effects of triglycerides and glucose, but not TyG.

4.3 External (temporal) validation and transportability

Both time-matched external AUROC of 0.652 (10-year censored) and unrestricted AUROC of 0.597 (~20-year follow-up) represent two phenomena. The decrease from 0.853 (development) to 0.652 (time-matched external) represents true temporal transportability loss, a typical decrease of 0.15–0.25 reported in published literature on cardiovascular external validation of tests [16,18]. Unrestricted follow-up brings the value down to 0.597, which is more or less an artifact of horizon mismatch, because both the model is trained with an 8.2-year median in the time window and is evaluated in 17 years. In Supplementary Table S8, an external characterisation by the different demographic subgroups is provided.

More importantly, the TyG dose–response gradient was not lost due to temporal transport or horizon mismatch. This dissociation - a preserved biological gradient - a degraded binary-rank discrimination, is a 'hallmark' of a valid generalizing effect. If TyG was just part of a bobsled team from 2003 to 2018, the gradient would have been flat in 1999-2002, which it was not.

Recalibration of Platt resulted in acceptable calibration (Hosmer–Lemeshow, $p = 0.164$) and normalized the EO ratio to 1.00, while the Brier score dropped from 0.101 to 0.081. Intercept and logistic recalibration resulted in mean biases, which were corrected, but miscalibrated group biases were still found. Platt scaling is our suggested approach for planting in new populations.

Some additional context from the perspective of the HOMA-IR comparison: The discrimination was similar between TyG and HOMA-IR (AUROC 0.590 vs 0.586; $r = 0.562$), and the dose–response gradients were comparable. This indicates that these two point to a similar underlying IR construct, with similar SNR relative to the prediction of CVD mortality. TyG is the costly but sensible panel for a resource-constrained environment, as its testing cost is ~20 times that of the fasting-glucose panel.

4.4 Clinical Implications

This study will only emphasize three clinically relevant items: first, that some families do not remember what is meant by "fasting" blood sugar; second, that the level of fasting insulin does not increase with age in the normal range; and third, that in most of Iraq, much of the Middle East, sub-Saharan Africa, and in many rural regions of high income countries, doctors do not have access to fasting insulin levels.

Pragmatic use of TyG: AUROC for discriminating the mortality of CVD is 0.590 and 0.586 for TyG and HOMA-IR, respectively. Clinically this is relevant as TyG requires a blood sugar level sample taken around 12 hours after the last meal (fasting glucose sample) plus lipids (triglycerides), which can be done in almost all clinical laboratories, whereas HOMA-IR

requires a sample of blood which is taken after the patient has not eaten for two hours or more, (fasting) plus measurement of the insulin, often difficult and expensive to obtain in a clinic.

Predictive power: The logistic regression model has a negative predictive value of 99.5%, a good rule-out test for short-term mortality of CVD. Biological gradient: This study maintained a well-defined dose–response relationship (7.4% to 13.0% mortality from CVD) across eras. This linear increase in univariate risk indicates that TyG adds value not just when used as a tool for stratification at the population level or when counselling patients, but also in the absence of ML deployment, supporting the project's overall objective of providing easy access to a biologically validated, disease-specific, and cost-effective risk stratification tool.

There's another caution for the ML developers: They should not assume that adding mathematical composite indexes will have any impact on model performance if the component of the indexes is already in the model. A similar composite index (TyG-BMI, TyG-WC, TG/HDL ratio, non-HDL cholesterol) is also hypothesized to be applicable to the context of flexible learners when applied to a context other than TyG.

## 5. LIMITATIONS

These results have two caveats.

- Same-survey temporal validation. This also is not external validation, but rather temporal — for the cohort of 1999-2002, the NHANES structure shares infrastructure with the development sample. To provide generalizability, replication within geographically and ethnically diverse cohorts (e.g., ARIC, Kailuan, CHARLS, UK Biobank) is required.

- Age dominance. Age discrimination is the dominant form of discrimination overall, in line with the biology of CVD mortality. The full-cohort LR model is not trained among the age 35–64 stratum and applied to individuals with age 40–79, and achieved a lower AUROC of 0.530 (n = 9,563, 246 events), compared to a trained application to the @3.5 age range (35–64), which has an AUROC of 0.843 (n = 1,435). This first relates to between-age contrast, the latter to a dedicated within-range model.

## 6. CONCLUSIONS

This study has scientific rigor and offers a high degree of clinical utility in that it satisfactorily addresses the question that has not been asked in the development of predictive modeling: What is the incremental value of a derived composite score relative to the individual raw scores? Using NHANES 2003–2018 (n = 16,322) with external validation on NHANES 1999–2002 (n = 4,318), we demonstrate the following:

TyG index does not have additional discriminative power beyond its individual features (ΔAUROC −0.0002, 95% CI −0.004 to +0.003) when incorporated into multivariate model-logistic (ML) models.

Without covariate adjustment, logistic regression gives a good AUROC and Cox C-index of discrimination (0.853 and 0.888, respectively) and a clear TyG dose–response gradient (Q1: 0.91% to Q4: 3.61%, $p < 0.001$), which is reduced after covariate adjustment.

A biological gradient is found across an independent temporal cohort (7.4% → 13.0%, $p < 0.001$), and external AUROCs are time-matched (0.652 LR; 0.885 Uno's C-index Cox, 10 years).

The difference in AUROC is not significant between HOMA-IR and TyG as IR surrogates, but the use of TyG is easier since it does not require any measurement of insulin.

A recalibration that is recommended for deployment of the frozen model to new populations is Platt scaling: Brier 0.081, E:O 1.00, H-L $p = 0.164$.

Competing-risk analysis validates the specificity of CVD, which is reflected by the TyG HR (1.203 for CVD vs 1.004 for non-CVD mortality), suggesting that TyG is a cardiovascular-specific, rather than mortality, indicator.

In summary, these results point to the proper use of TyG in clinical practice and model development for machine learning with multivariate models: it is well suited for manual use as an excellent screening/staging option, but it is a redundant feature for multivariate machine learning models which already include triglycerides and fasting glucose.

Natural Resources for Human Health
eISSN: 2583-1194

**Abbreviations**

| | |
|---|---|
| **AIC** | **Akaike Information Criterion** |
| **AUROC** | Area Under the Receiver Operating Characteristic Curve |
| **AUPRC** | Area Under the Precision-Recall Curve |
| **BMI** | Body Mass Index |
| **CDC** | Centers for Disease Control and Prevention |
| **CI** | Confidence Interval |
| **CVD** | Cardiovascular Disease |
| **Cox PH** | Cox Proportional Hazards |
| **DCA** | Decision Curve Analysis |
| **E:O** | Expected-to-Observed ratio |
| **FBG** | Fasting Blood Glucose |
| **HbA1c** | Glycated Hemoglobin |
| **HDL** | High-Density Lipoprotein |
| **H-L** | Hosmer–Lemeshow |
| **HOMA-IR** | Homeostasis Model Assessment of Insulin Resistance |
| **HR** | Hazard Ratio |
| **ICD-10** | International Classification of Diseases, 10th Revision |
| **IPCW** | Inverse Probability of Censoring Weighting |
| **IR** | Insulin Resistance |
| **LASSO** | Least Absolute Shrinkage and Selection Operator |
| **LDL** | Low-Density Lipoprotein |
| **LR** | Logistic Regression |
| **MI** | Myocardial Infarction |
| **ML** | Machine Learning |
| **NCHS** | National Center for Health Statistics |
| **NDI** | National Death Index |
| **NHANES** | National Health and Nutrition Examination Survey |
| **NPV** | Negative Predictive Value |
| **NRI** | Net Reclassification Index |
| **PCE** | Pooled Cohort Equations |
| **PPV** | Positive Predictive Value |
| **RF** | Random Forest |
| **SD** | Standard Deviation |

| | |
|---|---|
| **SHAP** | SHapley Additive exPlanations |
| **SRL** | Self-Reflective Learning |
| **TG** | Triglycerides |
| **TyG** | Triglyceride-Glucose index |
| **TyG-BMI** | Triglyceride-Glucose Body Mass Index |
| **WC** | Waist Circumference |
| **XGBoost** | Extreme Gradient Boosting |

Author Contributions:

S.N participated in all work packages from designing the study to machine learning pipeline implementation, data analysis, and the Drafting of the manuscript. W.M.A., A.A.E., H.K.S., B.K.M., and R.J.A. gave clinical oversight, medical interpretation, outcome definition, and subgroup review and provided manuscript review. Drafters of the paper (all the authors) approved it and agreed for final version.

Conflicts of Interest:

The authors do not have any conflicts of interest to disclose.

Funding:

There were no 'outside' funding sources.

Ethics Approval:

The study has been approved by the National Center for Training and Human Development in Kirkuk Directorate of Health/Iraqi Ministry of Health (11 May 2026, Approval No. 449). The information for this study is de-identified and publically available from NHANES. All data collection was approved by the NHANES IRB, and all of the participants signed an informed consent form during data collection.

Generative AI (Claude, Anthropic, Claude Opus 4, 2026) uses included in manuscript language polishing, formatting, and generation of a DOCX document from structured notes. The first author (SN) ran the ML pipeline and carried out all the statistical analysis, data interpretation, and scientific conclusions, which were checked by the co-author (WMA). All the content presented has been confirmed by both authors, who assume 100% of responsibility for its content.

Data Availability

The NHANES are made public at http://www.cdc.gov/nchs/nhanes. Mortality files linked to the NDI can be found at: https://www.cdc.gov/nchs/data-linkage/mortality.htm.

**REFERENCES**

[1] Simental-Mendía LE, Rodríguez-Morán M, Guerrero-Romero F. The product of fasting glucose and triglycerides as a surrogate for identifying insulin resistance in apparently healthy subjects. Metab Syndr Relat Disord. 2008;6 (4):299-304.
[2] da Silva A, Caldas APS, Hermsdorff HHM, et al. Triglyceride-glucose index is associated with symptomatic coronary artery disease in patients in secondary care. Cardiovasc Diabetol. 2019;18 (1):89.
[3] Tao LC, Xu JN, Wang TT, Hua F, Li JJ. Triglyceride-glucose index as a marker in cardiovascular diseases: landscape and limitations. Cardiovasc Diabetol. 2022;21 (1):68.
[4] Park GM, Cho YR, Won KB, et al. Triglyceride-glucose index is a useful marker for predicting subclinical coronary artery disease. Heart. 2021;107 (9):729-735.
[5] Wang X, Liu J, Wang D, et al. Association of the triglyceride-glucose index with all-cause and cardiovascular mortality among the general US adult population: NHANES 1999-2018. Cardiovasc Diabetol. 2022;21 (1):222.

[6] Li W, Shen C, Kong W, et al. Association between the triglyceride glucose index and all-cause and CVD mortality in the young population with diabetes. Cardiovasc Diabetol. 2023;22 (1):133.
[7] D'Agostino RB Sr, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation. 2008;117 (6):743-753.
[8] SCORE2 Working Group and ESC Cardiovascular Risk Collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Heart J. 2021;42 (25):2439-2454.
[9] Walonoski J, Kramer M, Nichols J, et al. Synthea: an approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. J Am Med Inform Assoc. 2018;25 (3):230-238.
[10] National Center for Health Statistics. NCHS data linked to NDI mortality files. https://www.cdc.gov/nchs/data-linkage/mortality.htm. Accessed 2026.
[11] Goff DC Jr, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129 (25 Suppl 2): S49-73.
[12] **12.** Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. Epidemiology. 2014;25 (1):114-121.
[13] Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. 2nd ed. Springer; 2019.
[14] Debray TPA, Vergouwe Y, Koffijberg H, et al. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol. 2015;68 (3):279-289.
[15] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35 (29):1925-1931.
[16] Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17 (1):230.
[17] Collins GS, Moons KGM, 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.
[18] Damen JA, Pajouheshnia R, Heus P, et al. Performance of the Framingham risk models and Pooled Cohort Equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis. BMC Med. 2019;17 (1):109.
[19] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26 (6):565-574.
[20] Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med. 2011;30 (10):1105-1117.
[21] World Health Organization. Cardiovascular diseases (CVDs) fact sheet. Geneva: WHO; 2023. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).