Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Adjustment choices shape apparent transportability of prognostic indices: the role of cohort-specific confounders across MIMIC-IV, eICU-CRD and CHARLS

Ruijie Lin
Wenzhou Medical University
Yanche Lu
Wenzhou Medical University
Chengze Ye
Hangzhou Normal University
Qiuyu Zheng
The First Affiliated Hospital of Wenzhou Medical University
Qian Yu
Wenzhou Medical University
Xueshu Yu
The First Affiliated Hospital of Wenzhou Medical University
Lianmin Ye
ylmwork88@126.com

The First Affiliated Hospital of Wenzhou Medical University

Research Article

Keywords: insulin resistance, triglyceride-glucose index, TyHGB, transportability, heterogeneity, MIMIC-IV, CHARLS

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11308952/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background. Insulin-resistance (IR) surrogate indices, including the triglyceride-glucose (TyG) index, are increasingly proposed as prognostic markers. Their number is expanding rapidly, yet it is unclear whether the prognostic information they carry is transportable between care settings. We examined how omitting cohort-specific confounders from the common adjustment set alters the apparent transportability of four IR surrogate indices across care settings.

Methods. Multi-database cohort study in three populations: MIMIC-IV (n = 1934; 28-day ICU mortality), eICU-CRD (n = 17262; in-hospital mortality), and CHARLS 2011 (n = 9324; mortality to 2018; 2015 wave for fasting sensitivity). ICU analyses used a 24-hour landmark. TyG, TyHGB, NHHR and TCBI used identical formulas (Table S1). Cohort-specific Cox models with three adjustment levels were fitted; per-SD hazard ratios were pooled by inverse-variance meta-analysis with I^2 and 95% prediction intervals.

Results. TyG had a positive association estimate in every cohort (HR 1.074–1.083) and TyHGB in all three (1.065–1.128; pooled 1.078, 95% CI 1.040–1.117). Primary cohort-specific adjustment yielded I^2 = 0% for all indices; common adjustment suggested I^2 of 70.2% (TyG) and 84.5% (TyHGB) (95% CIs 0–99.2% and 38.7–99.6%; overlapping with the 0% under cohort-specific adjustment). NHHR and TCBI showed I^2 = 0% under both strategies. Prediction intervals for TyG (0.828–1.403) and TyHGB (0.854–1.361) included unity. ΔAUC after adding individual indices ranged from − 0.0043 to 0.0053; inverse-probability weighting did not support reading unweighted ICU estimates as upper bounds.

Conclusions. Under per-SD scaling, cohort-specific adjustment yielded I^2 = 0% for all four indices, whereas common adjustment suggested I^2 of 70.2% (TyG) and 84.5% (TyHGB), though 95% CIs were wide and overlapping under k = 3. Reintroducing cohort-specific confounders (SOFA, APACHE IV) returned I^2 to 0%, consistent with residual confounding from omitting severity measures rather than a general adjustment-choice effect. Pooled hazard ratios combine different mortality horizons and do not constitute clinical transfer. Assessment of candidate indices should distinguish measurement calibration, conditional association, fixed-rule performance and incremental prediction for a defined population and horizon.

1. Introduction

Prognostic indices are increasingly proposed for use beyond the population in which they were derived, yet the evidence supporting such transfers rests on summaries that are themselves sensitive to analytic choices. The same index can yield a positive, null or heterogeneous pooled estimate depending on which covariates are adjusted for, on the exposure scale, and on the outcome horizon chosen. Between-cohort heterogeneity is commonly read as a sign that an association does not transport, but I^2 is a property of the fitted models as much as of the populations compared, and it is best interpreted with a confidence interval and alongside a prediction interval, as has repeatedly been recommended.[28,29] Whether the

apparent transportability of a prognostic index is a property of the index or of the analysis has not, to our knowledge, been examined within one set of participants under controlled changes of specification.

We address this question with a deliberately demanding empirical case, insulin-resistance (IR) surrogate indices. Direct measurement of insulin resistance is impractical at scale, so a large family of indices computed from routine laboratory values has been developed since the triglyceride-glucose (TyG) index was proposed in 2008,[1] followed by indices combining lipid fractions, glycaemia and adiposity.[2,3] New proposals continue to appear, including modified lipid ratios,[4] glucose-disposal proxies[5] and composite inflammatory or adiposity indices,[6–8] typically validated in a single cohort against a single outcome such as stroke,[9] cardiovascular disease in dysglycaemia,[10] cardiometabolic multimorbidity,[11] or mortality in critically ill[12] and community populations.[13] Using four indices that span the main construction strategies — TyG, TyHGB, NHHR and TCBI — with identical formulas and plausibility ranges, we examined how the established summaries of their prognostic performance changed with the adjustment set, the exposure scale and the outcome horizon, and we assessed fixed-rule performance and incremental prediction as separate questions.

2. Methods

2.1 Design and data sources

Multi-database cohort study in four populations; reporting follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement[14] and, for the secondary prediction analyses, the TRIPOD + AI statement.[15] Data sources were MIMIC-IV (v3.1, single-centre US electronic health records),[16] eICU-CRD (v2.0, multi-centre US tele-ICU)[17] and CHARLS (nationally representative Chinese adults aged ≥ 45 years),[18] with outcomes of 28-day ICU mortality, in-hospital mortality and all-cause mortality to 2018.

2.2 Populations

For MIMIC-IV we retained the first ICU admission per patient and required triglycerides, total cholesterol, HDL-cholesterol and glucose within 24 hours either side of ICU admission, with available body mass index (BMI). Adults with recorded sex and known outcomes entered ICU analyses if alive and still hospitalised at 24 hours; CHARLS participants were included if aged ≥ 45 years with a baseline 2011 blood sample, positive follow-up and computable indices, and the 2015 wave was used only for the fasting sensitivity analysis. Sample sizes, per-index complete cases and the covariate-complete target populations used in the selection analysis are given in Table S2 [see Additional file 1]; a flow diagram is Fig. 1.

2.3 Exposures

Four indices were computed with identical formulas in every database:

TyG = ln [TG (mg/dL) × glucose (mg/dL) / 2]

TyHGB = TG (mmol/L) / HDL-C (mmol/L) + 0.7 × glucose (mmol/L) + 0.1 × BMI (kg/m^2)[19]

NHHR = (total cholesterol − HDL-C) / HDL-C

TCBI = total cholesterol (mg/dL) × TG (mg/dL) × weight (kg) / 1000

Identical plausibility windows were applied throughout (full list in Additional file 2). In the ICU databases the first valid measurement within the window was used; in CHARLS the 2011 baseline sample. MIMIC height was the stay-level median and weight the admission weight; eICU used admission anthropometry. Where concentrations are reported, SI equivalents apply (glucose, mg/dL / 18.0182 = mmol/L; triglycerides, mg/dL × 0.0113 = mmol/L; total and HDL cholesterol, mg/dL × 0.0259 = mmol/L).

2.4 Outcomes and covariates

Primary outcomes were mortality from the 24-hour landmark to ICU-admission day 28 (MIMIC-IV), subsequent in-hospital mortality (eICU-CRD) and all-cause mortality over a median of 7 years (CHARLS). Prediction and additive-interaction analyses used horizons of 27 days after the MIMIC landmark, seven days after the eICU landmark and five years in CHARLS. Three nested models were fitted: Model 1 unadjusted; Model 2 plus age and sex; Model 3 plus BMI, diabetes and cohort-specific comorbidity and severity (SOFA and Charlson index in MIMIC-IV;[20] APACHE IV and a comorbidity count in eICU-CRD; smoking, hypertension and a chronic-condition count in CHARLS). Definitions and ascertainment sources are given in Additional file 2; missing covariates were retained as missing.

2.5 Statistical analysis

Primary analysis. Cox models were fitted with each index as a continuous variable standardised to its cohort-specific standard deviation (per-SD hazard ratio), and cohort-specific log hazard ratios were pooled by inverse-variance meta-analysis using DerSimonian–Laird random-effects estimates. The DerSimonian–Laird τ^2 moment estimator and I^2 and I^2 point estimates were computed by a hand-implemented formula (analyse.R, lines 22–25); the I^2 confidence intervals and the REML/HKSJ sensitivity analysis (Table S21) were computed with the metafor package, and the 95% confidence intervals for I^2 were derived using the Q-profile method (metafor::confint, type = 'QP'). Prediction intervals used a t-distribution multiplier with k − 2 degrees of freedom. REML and Hartung–Knapp–Sidik–Jonkman (HKSJ) adjusted estimates are reported in Table S21 as a sensitivity analysis. With k = 3 all $\tau 2$ estimators are imprecise, so $\tau 2$, Cochran's Q and 95% prediction intervals are reported alongside I2.21 Proportional-hazards departures in MIMIC-IV (Table S9) mean the corresponding hazard ratios are time-averaged.

I^2 served as a descriptive characterisation of between-cohort variation, not a decision threshold for clinical transportability. We use 'apparent transportability' for the observation of low between-cohort

heterogeneity in pooled estimates, a necessary but not sufficient condition for clinical transportability. Because I^2 is imprecise with few cohorts, we additionally report τ^2, Q and 95% prediction intervals.

With k = 3, the 95% confidence intervals for I^2 are wide and overlap between the two adjustment strategies, so the shift from 0% to 70% should be read as a qualitative pattern rather than a precise estimate. These cohorts do not jointly provide the comparable populations, measurement, adjustment and outcome horizons that exchangeability would require, so pooled estimates do not estimate a single transportable association. Prediction intervals used the random-effects mean and variance with a t multiplier (one degree of freedom).

Secondary analyses (detailed in Additional file 2). Dose–response was examined by quartiles and four-knot restricted cubic splines;[22] scaling was repeated with interquartile range and log transforms. ICU-derived Youden cut-offs were applied unchanged to CHARLS using censoring-weighted classification performance. Incremental value was assessed by AUC, IDI, continuous NRI[23] and decision curve analysis;[24] additive interaction by RERI;[25] and gradient-boosted trees with SHAP interpretation described fitted relationships.[26–27] Prediction used five held-out folds; Model 3 remained primary.

Sensitivity analyses covered documented fasting status in CHARLS 2015, CHARLS death-interval endpoints and 2018-wave known status, inverse-probability-of-selection weighting, and a common adjustment set (age, sex, BMI, diabetes, hypertension) on matched samples. Two variants probed the source of heterogeneity: the common set was extended with a harmonised severity proxy, and the Model 3 set for TyHGB was refitted with BMI removed. Both were specified after the primary analysis and are exploratory. Pre-specification. The four-index comparison, the choice of between-cohort heterogeneity as the primary criterion and the three-level modelling strategy were fixed before analysis; the plan was not registered, and stratified interaction results are hypothesis-generating with no multiplicity correction. Software. R 4.6.0 and Python 3.12 (packages in Additional file 2). Generative AI (OpenAI ChatGPT/Codex) was used to assist language revision, analysis-code development and figure preparation; all outputs were verified by the authors, who take full responsibility for the content.

3. Results

We report index-specific associations and how analysis choices alter their synthesis, followed by fixed-threshold and prediction analyses.

3.1 Cohorts

The analytic populations comprised 1934 ICU admissions in MIMIC-IV (321 deaths; 16.6%), 17262 in eICU-CRD (1057 deaths; 6.1%) and 9324 community-dwelling Chinese adults in CHARLS (861 deaths; 9.2%) (Table 1, Fig. 1). ICU participants were older and heavier than those in CHARLS. TyG was stable across cohorts (mean 8.94, 9.02 and 8.69), whereas TyHGB differed substantially (10.63, 11.12 and 8.10), reflecting differences in its lipid, glucose and BMI components.

Table 1
Baseline characteristics of the three study cohorts

Characteristic	MIMIC-IV	eICU-CRD	CHARLS
Number	1934	17262	9324
Deaths, n (%)	321/1934 (16.6%)	1057/17262 (6.1%)	861/9324 (9.2%)
Age, years	66.67 (14.85)	64.17 (14.32)	59.59 (9.36)
Female, n (%)	756/1934 (39.1%)	7073/17262 (41.0%)	4968/9324 (53.3%)
BMI, kg/m^2	29.00 (6.99)	29.72 (7.69)	23.52 (3.92)
Glucose, mg/dL	149.44 (74.22)	160.26 (87.59)	110.50 (36.76)
TyG	8.94 (0.76)	9.02 (0.81)	8.69 (0.67)
TyHGB	10.63 (5.08)	11.12 (5.09)	8.10 (3.21)
NHHR	2.92 (2.31)	3.18 (2.18)	3.11 (1.60)
TCBI	2218.89 (5498.74)	2203.78 (3586.85)	1650.31 (2026.86)
Diabetes, n (%)	592/1934 (30.6%)	2231/17262 (12.9%)	571/9221 (6.2%)
SOFA / APACHE IV	4.01 (3.39)	49.31 (22.65)	Not applicable
Comorbidity score/count	5.43 (2.69)	0.47 (0.75)	0.47 (0.72)

Full machine-readable version [see Additional file 8]. Values are mean (SD) or *n*/observed (%).

Footnote: death counts refer to the cohort with all four indices available; model-specific sample sizes and event counts, which vary with covariate completeness, are given in Table 2. MIMIC-IV and eICU-CRD observations begin at the 24-hour ICU landmark. CHARLS follow-up begins at the 2011 interview. Denominators are shown for categorical variables with missing data; APACHE IV was available for 15025 eICU participants and the comorbidity count for 9152 CHARLS participants.

3.2 Main associations

In fully adjusted models TyG had a positive association estimate in every cohort: HR per SD 1.081 (95% CI 0.994–1.175, P = 0.069) in MIMIC-IV, 1.074 (1.012–1.139, P = 0.018) in eICU-CRD and 1.083 (1.005–1.167, P = 0.038) in CHARLS (Table 2). TyHGB was positive in all three cohorts — 1.128 (1.037–1.227), 1.069 (1.011–1.130) and 1.065 (1.006–1.127) — and significant in each. NHHR and TCBI were not significantly associated with mortality, and the direction of the TCBI association differed between cohorts.

Table 2
Association of the four indices with mortality (hazard ratio per SD)

Cohort	Index	n / deaths	Model 1 HR (95% CI)	Model 2 HR (95% CI)	Model 3 HR (95% CI)	M3 P
MIMIC-IV	TyG	3114 / 616	1.118 (1.036–1.206)	1.221 (1.129–1.321)	1.081 (0.994–1.175)	0.069
MIMIC-IV	TyHGB	1951 / 323	1.117 (1.035–1.206)	1.182 (1.104–1.265)	1.128 (1.037–1.227)	0.005
MIMIC-IV	NHHR	1970 / 324	0.891 (0.776–1.023)	1.021 (0.901–1.157)	0.998 (0.884–1.127)	0.972
MIMIC-IV	TCBI	1979 / 332	0.875 (0.712–1.074)	0.991 (0.855–1.149)	0.986 (0.848–1.147)	0.860
eICU-CRD	TyG	17573 / 1360	1.110 (1.045–1.179)	1.197 (1.112–1.288)	1.074 (1.012–1.139)	0.018
eICU-CRD	TyHGB	15111 / 952	1.120 (1.081–1.159)	1.190 (1.142–1.240)	1.069 (1.011–1.130)	0.019
eICU-CRD	NHHR	15530 / 962	0.934 (0.873–1.000)	1.006 (0.945–1.072)	0.944 (0.862–1.035)	0.223
eICU-CRD	TCBI	15474 / 974	0.826 (0.717–0.951)	0.929 (0.829–1.040)	0.937 (0.846–1.037)	0.205
CHARLS	TyG	9135 / 842	0.982 (0.917–1.051)	1.074 (1.000–1.153)	1.083 (1.005–1.167)	0.038
CHARLS	TyHGB	9132 / 841	0.980 (0.911–1.054)	1.057 (0.996–1.121)	1.065 (1.006–1.127)	0.030
CHARLS	NHHR	9139 / 843	0.889 (0.822–0.961)	0.957 (0.885–1.034)	0.977 (0.901–1.059)	0.564
CHARLS	TCBI	9134 / 839	0.820 (0.733–0.918)	0.990 (0.905–1.083)	1.016 (0.933–1.107)	0.714

3.3 Transportability

With only k = 3 cohorts, no formal test of the difference in I^2 between adjustment strategies is available; the 95% confidence intervals for I^2 under the two strategies overlap substantially (TyG: 0–0% vs 0–99.2%; TyHGB: 0–97.4% vs 38.7–99.6%). The observed I^2 shift should therefore be interpreted as a qualitative pattern consistent with adjustment sensitivity rather than a precisely estimated contrast. Pooling the primary cohort-specific Model 3 estimates by inverse variance showed positive associations for TyG and TyHGB (Table 3, Fig. 2). TyG (1.078, 95% CI 1.035–1.123) and TyHGB (1.078, 1.040–1.117) had I2 = 0% under per-SD scaling ($\tau 2 = 0$; 95% CI for I2: TyG 0–0%, TyHGB 0–97.4%), whereas pooled estimates were close to the null for NHHR (0.969, 0.918–1.023) and TCBI (0.983, 0.925–1.044). On the same participants and SDs, common adjustment produced TyG HRs of 1.225/1.227/1.085 (I2 = 70.2%, 95% CI 0–99.2%) and TyHGB HRs of 1.193/1.208/1.063 (I2 = 84.5%, 95% CI 38.7–99.6%); fixed physical increments retained I2 = 0% (Tables S4 and S13 [see Additional file 1]; Figure S6 [see Additional file 3]), indicating that the per-SD agreement reflects scale-level rather than absolute-level consistency. Although the confidence intervals for TyG and TyHGB excluded unity, their 95% prediction intervals were wide and included unity (0.828–1.403 and 0.854–1.361), and with only three cohorts they are imprecise. NHHR and TCBI, which showed near-null associations in all cohorts, retained $I^2 = 0\%$ under both cohort-specific and common adjustment (Table 3), indicating that adjustment sensitivity was confined to indices with non-null associations — a pattern consistent with differential omitted-variable bias rather than a general effect of adjustment choice. Sensitivity re-estimation with restricted maximum likelihood (REML) and Hartung–Knapp–Sidik–Jonkman (HKSJ) adjusted confidence intervals (Table S21) confirmed the qualitative pattern. REML τ^2 estimates were highly consistent with the hand-written DerSimonian–Laird values (TyG: 0.00354 vs 0.00360; TyHGB: 0.00419 vs 0.00454), and REML I^2 was 69.9% (TyG) and 83.4% (TyHGB) under common adjustment, with $I^2 = 0\%$ under cohort-specific adjustment for all four indices. The HKSJ-adjusted 95% CIs were wider than the z-based intervals and crossed unity for both TyG (0.986–1.403) and TyHGB (0.967–1.374) under common adjustment, reinforcing the conclusion that with k = 3 cohorts only qualitative inference is warranted: the direction of the I^2 difference is stable, but its precise magnitude cannot be established.

Table 3
Descriptive pooled conditional associations and between-cohort heterogeneity

Index	Fixed effect HR (95% CI)	Random effects HR (95% CI)	95% prediction interval	I^2 (%)	τ^2	Q P
TyG	1.078 (1.035–1.123)	1.078 (1.035–1.123)	0.828–1.403	0.0	0.0000	0.983
TyHGB	1.078 (1.040–1.117)	1.078 (1.040–1.117)	0.854–1.361	0.0	0.0000	0.500
NHHR	0.969 (0.918–1.023)	0.969 (0.918–1.023)	0.682–1.378	0.0	0.0000	0.757
TCBI	0.983 (0.925–1.044)	0.983 (0.925–1.044)	0.666–1.451	0.0	0.0000	0.484

Full version [see Additional file 8].

Footnote: pooled HRs are descriptive syntheses of cohort-specific conditional associations with different mortality horizons and adjustment sets. The 95% prediction interval is a random-effects summary for an additional exchangeable study, not a calibrated forecast for a specified clinical task. With three cohorts the interval is wide (df = k − 2 = 1); every interval includes unity.

3.4 Secondary analyses (summary; details in Additional file 2)

All P values in this and subsequent secondary analyses (Sections 3.4–3.6) are exploratory and uncorrected for multiplicity. Dose–response was non-linear for TyHGB in all three cohorts and for TyG in two (Figs. 3–4, Table S3 [see Additional file 1]). Fixed ICU cut-offs had low CHARLS sensitivity (11.2%–50.8%), and source apparent AUCs exceeded target continuous AUCs (Table 4, Figure S4 [see Additional file 9]). Incremental discrimination was minimal: ΔAUC ranged from − 0.0043 to 0.0053 for individual indices and from 0.0014 to 0.0044 for all four jointly (Table S6 [see Additional file 1], Figure S1 [see Additional file 4]). Full prediction model performance by cohort, model and index is reported in Table S5 [see Additional file 1]. No additive interaction was detected, and XGBoost did not improve discrimination (Figures S2 and S3 [see Additional files 6 and 5]). Fasting-status and death-timing sensitivities appear in Tables S7–S8 and S11–S12 and Additional file 10.

Table 4
Apparent source performance and CHARLS evaluation of fixed ICU cut-offs

Source / index	Cut-off	AUC S/T	Youden S/T	Sensitivity, % S/T (target 95% CI)	Specificity, % S/T (target 95% CI)	Target PPV, % (95% CI)	Target NPV, % (95% CI)
MIMIC-IV TyG	9.254	0.541/0.490	0.079/0.020	42.7/18.9 (15.2–22.4)	65.3/83.1 (82.1–83.8)	5.2 (4.1–6.2)	95.4 (95.0–95.9)
eICU-CRD TyG	8.581	0.556/0.490	0.094/-0.010	78.4/50.8 (46.1–55.5)	31.0/48.3 (47.2–49.3)	4.6 (4.0–5.2)	95.2 (94.6–95.8)
MIMIC-IV TyHGB	10.632	0.571/0.477	0.129/0.019	43.7/11.2 (8.2–14.1)	69.2/90.7 (90.1–91.4)	5.6 (4.0–7.0)	95.4 (94.9–95.9)
eICU-CRD TyHGB	10.100	0.602/0.477	0.167/0.019	60.7/13.3 (10.0–16.7)	56.0/88.6 (87.9–89.3)	5.4 (4.0–6.9)	95.4 (95.0–95.8)

Full version [see Additional file 8]. S/T denotes source/target. ICU source cut-offs were fixed in CHARLS, where five-year mortality used midpoint-assigned death intervals and censoring weights; interval-bound sensitivities are in Table S18. Source outcomes were mortality to ICU-admission day 28 in MIMIC-IV and in-hospital death within seven days after the eICU landmark. Source *n*/deaths were 3114/616 (MIMIC-IV TyG), 17573/951 (eICU-CRD TyG), 1951/323 (MIMIC-IV TyHGB) and 15111/699 (eICU-CRD TyHGB); target n was 9135 for TyG and 9132 for TyHGB. Source estimates are apparent because the same records selected each cut-off; they are not internally validated. AUC describes the continuous index, while Youden, sensitivity and specificity correspond to the shown cut-off. Parenthesised intervals in table cells are target-only 95% CIs from 500 bootstrap resamples; PPV and negative predictive value (NPV) are target-only values. The four-index prediction analyses used a smaller common complete-case sample.

3.5 Selection into the analytic sample

Because only 3.1% (MIMIC-IV) and 11.9% (eICU-CRD) of landmark-eligible admissions had all four indices computable, selection was examined directly by inverse-probability-of-selection weighting against the covariate-complete target populations (Tables S14 and S15 [see Additional file 1]). After weighting, TyG hazard ratios were 1.114 (1.019–1.217) and 1.063 (0.998–1.132), and TyHGB estimates 1.085 (0.997–1.181) and 1.050 (0.980–1.125), for MIMIC-IV and eICU-CRD. Not all estimates moved towards the null, so the ICU effect sizes cannot be read as upper bounds. Re-pooling weighted hazard ratios (k = 2) yielded I^2 = 0.0% for TyG and TyHGB (Table S20), consistent with unweighted Model 3.

3.5.1 Selection intensity differs across the three cohorts

The proportion of the eligible population with all four indices computable ranged from 3.1% (MIMIC-IV) through 11.9% (eICU-CRD) to 82.7% (CHARLS); weighting targets are smaller because they require complete Model 3 covariates. Selection counts and weighting diagnostics are in Table S16 [see Additional file 1].

3.6 Interaction by severity or age (exploratory)

A consistent pattern was not supported across all three cohorts (Table 5, Fig. 5). The lower-severity or younger stratum showed a stronger point estimate for the index concerned in some comparisons, but ICU interaction P values were all ≥ 0.139 and the single CHARLS TyG interaction (P = 0.028) was one of 12 exploratory comparisons. This analysis was data-driven and unreplicated and is reported as hypothesis-generating.

Table 5
Interaction by severity or age (hazard ratio per SD, fully adjusted)

Cohort	Index	Stratum cut-off	Lower stratum HR (95% CI); *n*/deaths	Higher stratum HR (95% CI); *n*/deaths	Interaction P
MIMIC-IV	TyG	severity 4	1.197 (0.947–1.512) 1257/137	1.055 (0.963–1.156) 1857/479	0.823
MIMIC-IV	TyHGB	severity 3	1.537 (1.096–2.153) 829/82	1.096 (1.004–1.196) 1122/241	0.862
MIMIC-IV	NHHR	severity 3	0.963 (0.648–1.430) 839/82	1.007 (0.889–1.140) 1131/242	0.152
MIMIC-IV	TCBI	severity 3	0.906 (0.427–1.922) 832/82	0.992 (0.855–1.151) 1147/250	0.140
eICU-CRD	TyG	severity 47	1.102 (0.905–1.343) 8498/143	1.067 (1.006–1.130) 9075/1217	0.647
eICU-CRD	TyHGB	severity 45	1.148 (0.932–1.413) 7380/103	1.059 (1.002–1.119) 7731/849	0.547
eICU-CRD	NHHR	severity 45	0.971 (0.808–1.165) 7643/104	0.946 (0.867–1.033) 7887/858	0.521
eICU-CRD	TCBI	severity	1.002	0.926	0.947

Full version [see Additional file 8]. Strata use the index-specific Model 3 median SOFA, APACHE IV score or age. HRs retain the same cohort- and index-specific full-sample SD across strata; the interaction *P* value comes from the full-sample product-term model. Each stratum reports *n*/events. CHARLS TyG interaction *P* = 0.028 was one of 12 exploratory comparisons and should be interpreted descriptively rather than as a confirmatory test.

Additional files

Cohort	Index	Stratum cut-off	Lower stratum HR (95% CI); *n*/deaths	Higher stratum HR (95% CI); *n*/deaths	Interaction P
		45	(0.848–1.185)	(0.830–1.033)	
			7592/104	7882/870	
CHARLS	TyG	age	1.232	1.036	0.028
		59	(1.062–1.430)	(0.950–1.129)	
			4493/147	4642/695	
CHARLS	TyHGB	age	1.098	1.049	0.335
		59	(0.994–1.213)	(0.975–1.129)	
			4492/146	4640/695	
CHARLS	NHHR	age	0.937	0.982	0.887
		59	(0.784–1.120)	(0.897–1.075)	
			4497/146	4642/697	
CHARLS	TCBI	age	1.036	0.975	0.319
		59	(0.935–1.148)	(0.858–1.109)	
			4497/147	4637/692	
Full version [see Additional file 8]. Strata use the index-specific Model 3 median SOFA, APACHE IV score or age. HRs retain the same cohort- and index-specific full-sample SD across strata; the interaction *P* value comes from the full-sample product-term model. Each stratum reports *n*/events. CHARLS TyG interaction *P* = 0.028 was one of 12 exploratory comparisons and should be interpreted descriptively rather than as a confirmatory test.					
Additional files					

4. Discussion

4.1 Principal findings

On identical participants and exposure scales, changing the adjustment set suggested a rise in TyG and TyHGB I2 from 0% to 70.2% and 84.5%, the central methodological observation of this study: apparent agreement depends on which conditional association is estimated. However, the 95% confidence intervals for I2 under k = 3 are wide and overlap between the two strategies (TyG: 0–0% vs 0–99.2%; TyHGB: 0–97.4% vs 38.7–99.6%), so this shift should be read as a qualitative pattern rather than a precise estimate. The rise could in principle reflect a mechanical consequence of the shared adjustment set, which omits SOFA and APACHE IV — strong severity predictors present only in the ICU cohorts (Table 1) — so that removing them shifts ICU coefficients towards their crude estimates while CHARLS

coefficients barely move. Adding a harmonised severity proxy to the common set did not reduce heterogeneity; the proxy was a coarse, diagnosis-based measure that excludes physiological severity components and cannot substitute for SOFA or APACHE IV. A stronger harmonised severity score (e.g., an ICD-code-based comorbidity severity score) might be more informative but was not constructed in the present data, so differential de-adjustment could not be ruled out. We accordingly reframe our central claim: the two adjustment strategies estimate different conditional associations, and the apparent transportability of a prognostic index is contingent on model specification. Primary pooled HRs near 1.08 combine different adjustment sets and mortality horizons and do not estimate one shared clinical task. Prediction intervals included unity and incremental discrimination was small, and because the cohorts differ in outcome horizon, I2 across them conflates population transportability with outcome-specification transportability.21

4.2 Comparison with existing literature

Reports of new IR surrogate indices have grown, and many rely on single-cohort associations and conventional P values. Our paired analyses offer an explanation for a recognised inconsistency in that literature: reported per-SD hazard ratios for TyG range from about 1.0 to 1.5 across recent cohorts,[10–13] and the handling of illness severity and comorbidity can change both the estimates and their heterogeneity. TyHGB had previously been evaluated only in community cohorts[9]; its association was positive in all three cohorts, yet its prediction interval included unity, illustrating why single-cohort significance testing alone is an unreliable basis for judging a marker. Considered methodologically, three bodies of work bear on these findings. Guidance for systematic reviews of prognostic factor studies recommends pre-specifying the adjustment set and reporting the full model specification,[28] and our paired comparison shows why: on identical participants, changing the adjustment set alone moved I^2 from 0% to 70.2% and 84.5%. Work on quantifying heterogeneity has long emphasised that I^2 is imprecise with few studies and is not a threshold for transportability,[21] and that prediction intervals, not confidence intervals, describe the range of effects expected in a new setting;[29] we extend that argument by showing that the heterogeneity statistic is itself a function of model specification. Finally, the STROBE and TRIPOD + AI reporting frameworks,[14,15] applied here to the same study, leave unaddressed precisely the question our paired analysis targets.

4.3 Interpretation

The prediction intervals summarise uncertainty under the fitted synthesis and cannot establish clinical transportability across these differing tasks. The association was stronger in younger CHARLS participants, but formal ICU interaction tests did not support a general concentration of associations in lower-severity patients. Illness severity could act as confounder, mediator or collider; this design cannot distinguish these possibilities and no causal claim is made.

A further complication is that TyHGB includes BMI as a component and Model 3 adjusts for BMI in all cohorts, so the total association is partially absorbed by the adjustment covariate; because BMI distributions differ markedly across cohorts (29.0, 29.7 and 23.5 kg/m2), the degree of absorption varies

and introduces a source of cohort-dependent attenuation. A sensitivity model omitting BMI left the TyHGB association materially unchanged, with the largest attenuation observed in CHARLS, consistent with its lowest BMI distribution. ICU acute-phase lipid metabolism may distort TyG and TyHGB from their chronic insulin-resistance meaning, so the indices may measure different biological constructs in ICU and community settings.

What the data show is that the two adjustment strategies estimate different quantities, and that the choice of adjustment set changes both the magnitude of the estimates and their apparent transportability. Under fixed physical increments the pooled estimates also retained I^2 = 0% (TyG 1.100, 95% CI 1.044–1.159; TyHGB 1.017, 95% CI 1.009–1.025), indicating that the per-SD agreement reflects scale-level rather than absolute-level consistency; our primary estimates should therefore be read as prognostic associations conditional on illness severity. The common set omits severity (SOFA, APACHE IV), Charlson index and smoking, and reintroducing cohort-specific confounders returned I^2 to 0% (Table S19), indicating that the I^2 rise reflects residual confounding rather than a pure choice effect. The comparison remains exploratory and was not pre-specified.

The mechanism underlying the I² rise under common adjustment can be understood as omitted-variable bias. The public adjustment set (age, sex, BMI, diabetes, hypertension) excludes SOFA and APACHE IV — strong severity predictors available only in the ICU cohorts. Removing these covariates from the ICU models mechanically shifts the TyG and TyHGB coefficients towards their crude estimates (e.g., MIMIC TyG HR moves from 1.081 under Model 3 to 1.225 under common adjustment), while the CHARLS coefficient barely changes (1.083 to 1.085) because CHARLS never included these severity measures. This differential movement inflates between-cohort heterogeneity for indices with non-null associations. Reintroducing the cohort-specific confounders (Table S19, M3_shared) returned I² to 0% for all four indices. This is consistent with residual confounding from omitting severity measures, although M3_shared is numerically identical to the primary Model 3 specification (it is not an independent validation), and the harmonised severity proxy did not confirm a severity-specific mechanism. The NHHR and TCBI indices, which were near-null in all cohorts, showed I² = 0% under both strategies, consistent with the expectation that omitted-variable bias selectively moves non-zero estimates rather than null ones.

4.4 Implications

These results do not support promoting an IR index for mortality prognostication from single-cohort significance or low I^2 alone. Heterogeneity and prediction intervals are established summaries; this case shows why paired adjustment comparisons, a defined prediction task and fixed-rule (fixed-threshold) performance also matter. Source cut-offs had weak apparent discrimination, and their CHARLS performance depended on death-time assignment, so these data do not isolate loss caused by transfer. More broadly, assessment of a candidate index should distinguish measurement calibration and scaling, conditional association and model dependence, fixed-rule performance, and incremental prediction for a defined population and horizon. We propose that future transportability claims report (a) pre-registered

adjustment sets, (b) I^2 with 95% CI and prediction intervals, and (c) sensitivity to common and cohort-specific adjustment.

4.5 Strengths and limitations

The study compared four populations using common index definitions and examined scaling, fasting, death timing, adjustment and selection sensitivity. Adjustment sets were necessarily cohort-specific, and the common-adjustment analysis showed that this choice materially affects heterogeneity (Figure S6).

Time-varying associations and departures from proportional hazards in ICU models limit interpretation of a single HR as constant throughout follow-up. Comorbidity ascertainment differed between databases, and smoking was unavailable in eICU-CRD. CHARLS death times were incomplete, and interval assignments changed five-year case counts, although incremental AUC gains remained small. Lipid panels were ordered at clinician discretion, and selection weighting addresses measured selection only: in MIMIC-IV, TyG was missing in 89.6% of landmark-eligible admissions and BMI in 49.6%, so only 3.1% had all four indices. eICU records represent admissions rather than unique people, MIMIC stay-level height and diagnosis summaries may postdate the landmark, the prediction analyses used five-fold internal cross-validation only and do not constitute external validation in the TRIPOD + AI sense; the fixed-cutoff analyses apply ICU-derived thresholds to CHARLS, but the source cut-offs were derived on the same records and are not independently validated, so these results are internally validated comparisons rather than estimates for a deployable model, and CHARLS loss to follow-up may be informative. Furthermore, inverse-probability weighting can only address measured selection; unmeasured selection likely operates in the same direction, as clinicians may order lipid panels preferentially for more severely ill patients, which would inflate the observed ICU associations relative to the underlying population. This is a directional argument, not a quantified correction.

The analysis plan was not registered, and all sensitivity and subgroup analyses were exploratory with no multiplicity correction. With three cohorts, heterogeneity and prediction intervals were estimated imprecisely, the 95% confidence intervals for I^2 were wide and overlapping between adjustment strategies, and an I^2 point estimate of 0% does not preclude meaningful heterogeneity. This is an observational study; no causal inference is warranted.

5. Conclusions

For TyG and TyHGB, the direction of the fitted association was positive in all three cohorts, but on the same participants I^2 rose from 0% under per-SD cohort-specific adjustment to 70.2% and 84.5% under common adjustment, although the 95% confidence intervals for I^2 were wide and overlapping under $k = 3$. Their pooled HRs combine different mortality tasks and do not establish clinical transfer. The methodological lesson is to check how adjustment, measurement and outcome choices affect association summaries, and then to evaluate fixed rules and incremental prediction for a specified target.

Abbreviations

AUC: area under the curve

APACHE IV: Acute Physiology and Chronic Health Evaluation IV

BMI: body mass index

CI: confidence interval

CITI: Collaborative Institutional Training Initiative

CRediT: Contributor Roles Taxonomy

DOI: digital object identifier

HDL-C: high-density lipoprotein cholesterol

HR: hazard ratio

ICD: International Classification of Diseases

IDI: integrated discrimination improvement

IQR: interquartile range

IR: insulin resistance

I^2: I-squared (between-cohort heterogeneity)

NHHR: non-HDL-to-HDL cholesterol ratio

NPV: negative predictive value

NRI: net reclassification improvement

ORCID: Open Researcher and Contributor ID

PPV: positive predictive value

RERI: relative excess risk due to interaction

SD: standard deviation

SHAP: SHapley Additive exPlanations

SOFA: Sequential Organ Failure Assessment

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

TCBI: triglyceride-cholesterol-body weight index

TRIPOD+AI: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis + Artificial Intelligence

TyG: triglyceride-glucose index

TyHGB: triglyceride-high-density lipoprotein-glucose-body mass index

XGBoost: eXtreme Gradient Boosting

τ^2: tau-squared (between-study variance)

Declarations

Ethics approval and consent to participate.

This study used only publicly available, de-identified databases (MIMIC-IV, eICU-CRD) and a publicly released research dataset (CHARLS). MIMIC-IV and eICU-CRD were accessed after completion of the required Collaborative Institutional Training Initiative (CITI) data-or-privacy training and execution of the PhysioNet data use agreement; establishment of MIMIC-IV was approved by the Institutional Review Boards of the Beth Israel Deaconess Medical Center (2001-P-001699/14) and MIT. CHARLS was approved by the Biomedical Ethics Review Committee of Peking University (IRB00001052-11015), and all participants provided written informed consent. The research was conducted in accordance with the Declaration of Helsinki. No additional ethical approval was required for this secondary analysis.

Clinical trial number

Not applicable.

Consent for publication.

Not applicable.

Availability of data and materials.

The datasets supporting the conclusions of this article are available in the PhysioNet repository [16,17] and the CHARLS repository [18], under data-use agreements that prohibit redistribution of individual-level records. MIMIC-IV (v3.1, DOI: https://doi.org/10.13026/kpb9-mt58) and eICU-CRD (v2.0) are accessible

via PhysioNet after completion of CITI training and execution of a data-use agreement. CHARLS is available from its official website [30] after registration. Analysis code, aggregate results and an upstream field map are included within the article (and its additional file 8).

Competing interests.

The authors declare that they have no competing interests.

Funding.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contributions (CRediT).

R.L. and Y.L. conceived and designed the study, performed the statistical analyses, and wrote the main manuscript text. C.Y., Q.Z. and Q.Y. curated and investigated the data; Q.Y. additionally prepared the figures. X.Y. provided study resources. L.Y. supervised the project, administered it, and substantively revised the manuscript. All authors read and approved the final manuscript.

Declaration of generative AI use.

OpenAI ChatGPT/Codex was used to assist language revision, analysis-code development, figure preparation and consistency checks. The authors take responsibility for the content, analyses and interpretation of the manuscript.

Acknowledgements.

The authors acknowledge the MIMIC-IV, eICU-CRD and CHARLS teams and participants for making these data publicly available.

Authors' information.

Ruijie Lin (ORCID: 0009-0003-0312-9248); Yanche Lu (ORCID: 0009-0001-9490-7951); Lianmin Ye (ORCID: 0000-0002-8454-2362).

References

1. Simental-Mendía LE, Rodríguez-Morán M, Guerrero-Romero F. The product of fasting glucose and triglycerides as surrogate for identifying insulin resistance in apparently healthy subjects. Metab Syndr Relat Disord. 2008;6(4):299–304. 10.1089/met.2008.0034. . PMID 19067533.
2. Song S, Zhang Y, Qiao X, Duo Y, Xu J, Peng Z, et al. HOMA-IR as a risk factor of gestational diabetes mellitus and a novel simple surrogate index in early pregnancy. Int J Gynaecol Obstet. 2022;157(3):694–701. 10.1002/ijgo.13905. . PMID 34449903.
3. Lu Y, Jiang D, Zhang J, Wang C, Li J, Liang Y. Association of a novel nutritional index, the triglyceride-cholesterol-body weight index (TCBI), with incident mild cognitive impairment: CHARLS 2011–2018. Front Nutr. 2026;13:1834512. 10.3389/fnut.2026.1834512. PMID 42325519.
4. Paulmichl K, Hatunic M, Højlund K, Jotic A, Krebs M, Mitrakou A et al. Modification and validation of the triglyceride-to-HDL cholesterol ratio as a surrogate of insulin sensitivity in white juveniles and adults without diabetes mellitus: the Single Point Insulin Sensitivity Estimator (SPISE). Clin Chem. 2016;62(9):1211–9. 10.1373/clinchem.2016.257436. PMID 27471037.
5. Olson JC, Erbey JR, Williams KV, Becker DJ, Edmundowicz D, Kelsey SF, et al. Subclinical atherosclerosis and estimated glucose disposal rate as predictors of mortality in type 1 diabetes. Ann Epidemiol. 2002;12(5):331–7. 10.1016/S1047-2797(01)00269-1. . PMID 12062921.
6. Liu M, Xu N, Sha G, Ye Y, Yan K, Zhang K, et al. A novel metabolic inflammation index for cardiovascular prognosis: evidence from three international cohorts in middle-aged and older adults [Epub ahead of print]. Diabetes Obes Metab. 2026. 10.1111/dom.71213. . PMID 42681850.
7. Chen J, Sun Z, Qiu J, Zhou Z, Ma K, Liang Y, et al. Composite C-reactive protein-triglyceride-glucose adiposity indices and incident cardiovascular-liver-metabolic multimorbidity: a prospective cohort study. Cardiovasc Diabetol. 2026;25(1):264. 10.1186/s12933-026-03351-5. . PMID 42665793.
8. Niu Y, Chen F, Wu R, Liu M, Deng S, Wang C, et al. Association of the CHG index combined with obesity indices and incident cardiometabolic multimorbidity in a nationwide prospective cohort study [Epub ahead of print]. Sci Rep. 2026. 10.1038/s41598-026-61144-y. . PMID 42414530.
9. Lu AM, Xu Q. Association between the triglyceride-high-density lipoprotein-glucose-body index and new-onset stroke risk: a prospective cohort study. Lipids Health Dis. 2026;25(1):143. 10.1186/s12944-026-02953-y. . PMID 42002766.
10. Xie H, Yan C, Zheng Y, Wu H. Assessment of ten insulin resistance surrogate indexes predicts new-onset cardiovascular disease incidence in patients with prediabetes or diabetes: insights from CHARLS data with machine learning analysis. Glob Heart. 2026;21(1):17. 10.5334/gh.1532. . PMID 41836049.
11. Zhao E, Xie H, Wang R, Dong B, Zhang C. Estimated glucose disposal rate outperforms thirteen other insulin resistance indices in predicting new-onset cardiometabolic multimorbidity: a longitudinal study from the China Health and Retirement Longitudinal Study (CHARLS) and cross-sectional replication from the National Health and Nutrition Examination Survey (NHANES). Cardiovasc Diabetol. 2026;25(1):205. 10.1186/s12933-026-03218-9. . PMID 42226282.

12. Wang X, Liu Y, Zhang Q, Xin Y, Shen L, Xiao Y et al. Triglyceride-glucose frailty index, metabolic-frailty phenotypes, and mortality in critically ill patients with acute kidney injury: derivation, interpretation, and external validation. Exp Gerontol. 2026;224:113291. 10.1016/j.exger.2026.113291. PMID 42632433.
13. Tang Y, Xu Y, Yue Z, Wang W, Wen S, Wang H, et al. Nine anthropometric insulin resistance indices predict mortality and life expectancy across glucose states. iScience. 2026;29(9):117400. 10.1016/j.isci.2026.117400. . PMID 42750797.
14. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453–7. 10.1016/S0140-6736(07)61602-X. . PMID 18064739.
15. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, 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. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10(1):1. 10.1038/s41597-022-01899-x. . PMID 36596836.
17. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data. 2018;5:180178. 10.1038/sdata.2018.178. PMID 30204154.
18. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol. 2014;43(1):61–8. 10.1093/ije/dys203. . PMID 23243115.
19. Xuan Y, Yu J, Yao Q. Waist circumference partially mediates the non-linear association between TyHGB and MASLD in non-diabetic Japanese adults. Front Nutr. 2026;13:1915937. 10.3389/fnut.2026.1915937.
20. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373–83. 10.1016/0021-9681(87)90171-8. . PMID 3558716.
21. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539–58. 10.1002/sim.1186. . PMID 12111919.
22. Desquilbet L, Mariotti F. Dose-response analyses using restricted cubic spline functions in public health research. Stat Med. 2010;29(9):1037–57. 10.1002/sim.3841. . PMID 20087875.
23. Pencina MJ, D'Agostino RB, Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008;27(2):157–72. 10.1002/sim.2929. . PMID 17569110.
24. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26(6):565–74. 10.1177/0272989X06295361. . PMID 17099194.

25. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol. 2012;41(2):514–20. 10.1093/ije/dyr218. . PMID 22253321.
26. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM; 2016. pp. 785–94. 10.1145/2939672.2939785
27. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 30 (NIPS 2017). Red Hook, NY: Curran Associates; 2017. pp. 4765–74.
28. Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ. 2019;364:k4597. 10.1136/bmj.k4597.
29. Graham PL, Moran JL. Robust meta-analytic conclusions mandate the provision of prediction intervals in meta-analysis summaries. J Clin Epidemiol. 2012;65(5):503–10. 10.1016/j.jclinepi.2011.09.012.
30. China Health and Retirement Longitudinal Study (CHARLS). Peking University. http://charls.pku.edu.cn/. Accessed 29 September 2026.

Figures

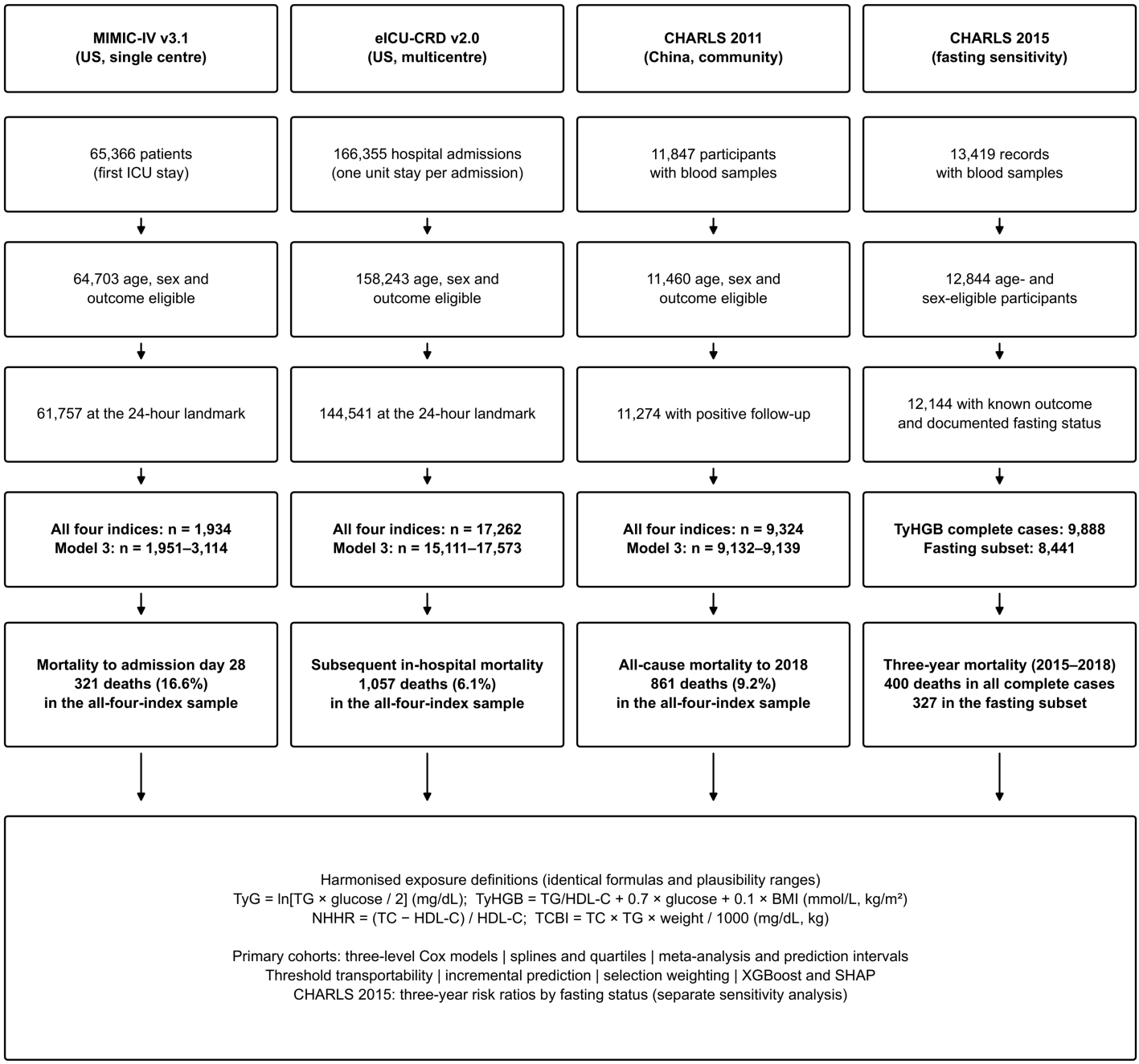


Figure 1

Study design and cohort selection. Four populations (MIMIC-IV, eICU-CRD, CHARLS 2011 and CHARLS 2015) with sequential inclusion steps, the resulting analytic samples, and the harmonised exposure definitions and analysis plan. Main descriptive samples have all four indices available; association models use index-specific complete cases. ICU follow-up begins at a 24-hour landmark. CHARLS 2015 is a separate fasting-status sensitivity analysis, not a fourth independent cohort in the pooled estimates. The CHARLS 2015 branch reports 12144 participants with known outcome and documented fasting status; the one-participant difference from 12145 reflects an additional record with a missing fasting flag.

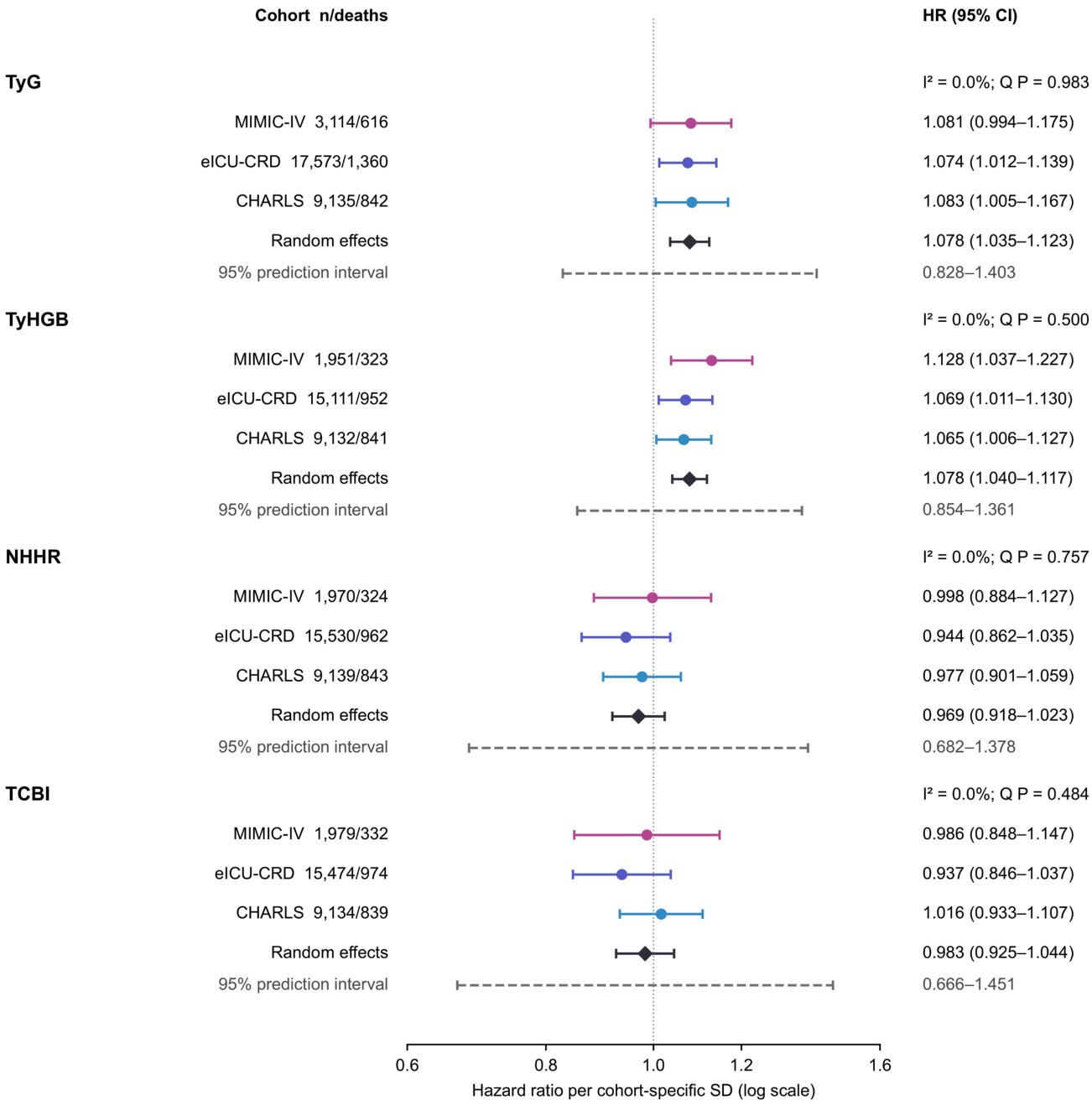


Figure 2

Associations of the four indices with mortality, by cohort and pooled. Forest plot of cohort-specific and pooled hazard ratios per SD with 95% confidence intervals from fully adjusted (Model 3) Cox models using index-specific Model 3 complete cases. Rows are grouped by index (TyG, TyHGB, NHHR, TCBI); within each group the three cohorts are shown with the number analysed and number of events, followed by the DerSimonian-Laird inverse-variance random-effects estimate (diamond) and 95% prediction interval. Between-cohort heterogeneity (I^2) and the Cochran Q-test *P* value are shown beside each index.

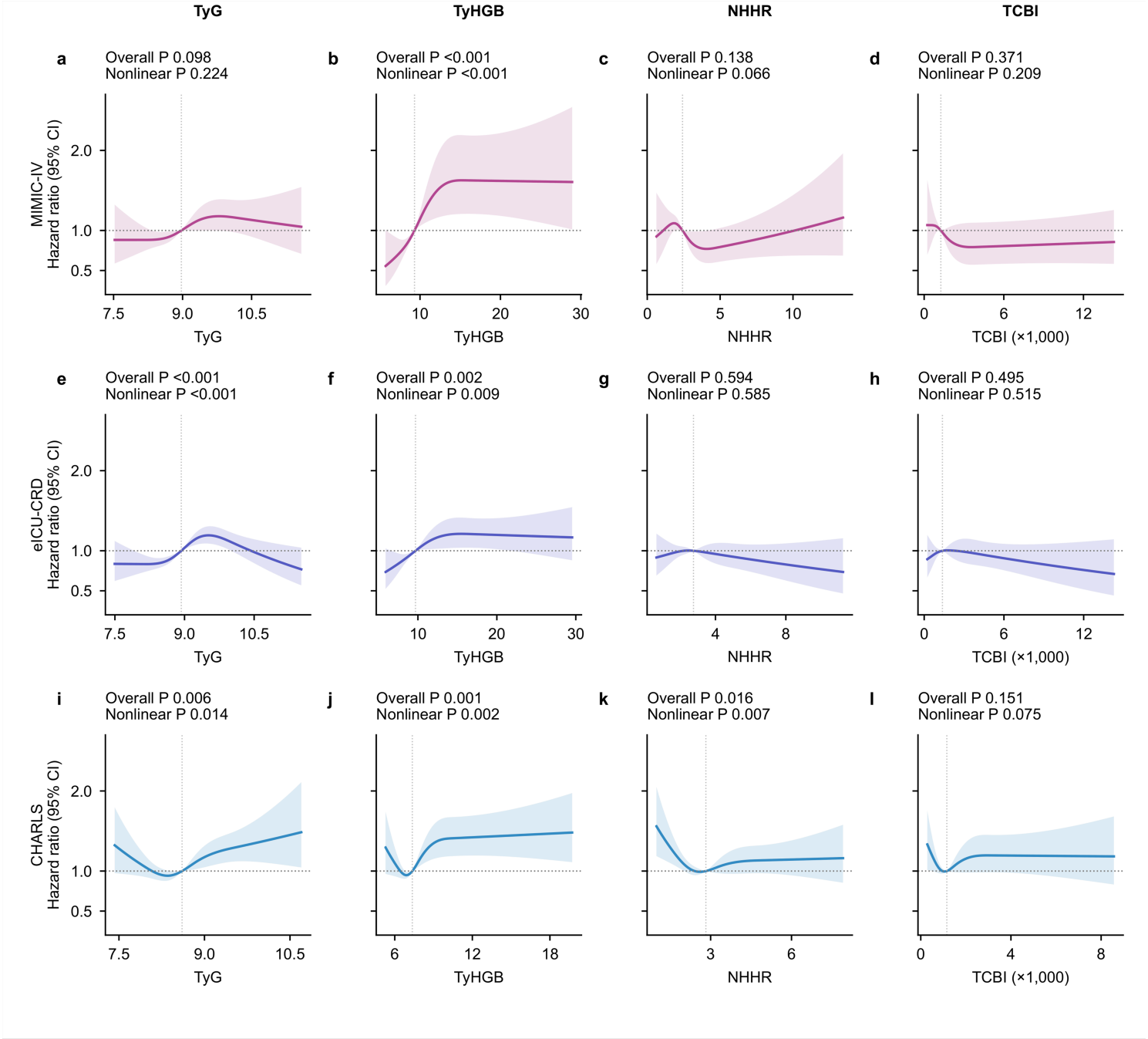


Figure 3

Dose–response relationships. Restricted cubic spline curves (four knots) for each index in each cohort, with hazard ratios referenced to the cohort-specific median and shaded 95% confidence bands; *P* values are shown for the overall association and for non-linearity. Panels are arranged as three rows by cohort (a–d, MIMIC-IV; e–h, eICU-CRD; i–l, CHARLS) and four columns by index (TyG, TyHGB, NHHR, TCBI). TCBI is shown in units of 1000.

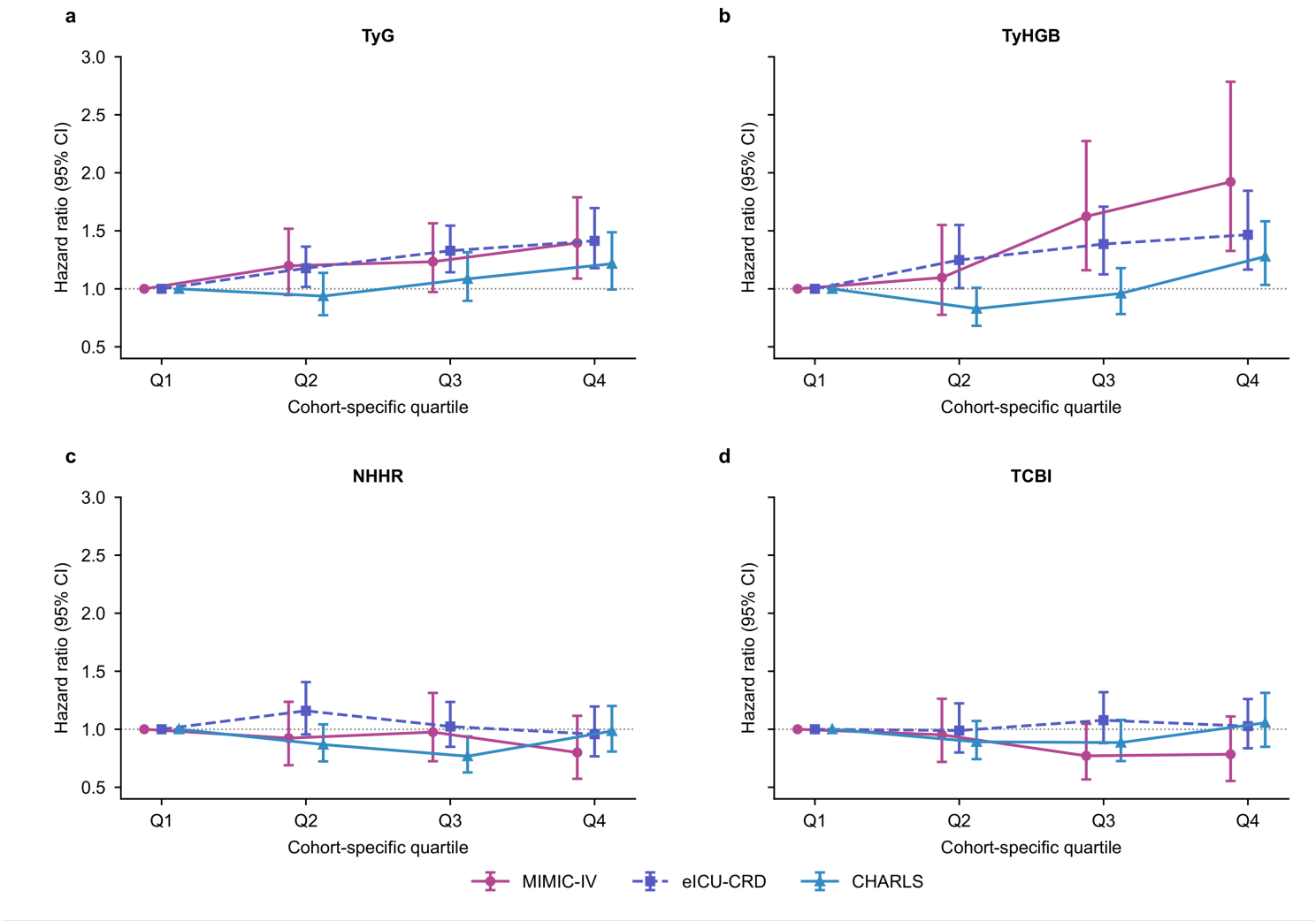


Figure 4

Quartile analysis. Adjusted hazard ratios with 95% confidence intervals across quartiles of each index (Q1 reference) in the three cohorts. Panels a–d correspond to TyG, TyHGB, NHHR and TCBI; within each panel the three cohorts are shown.

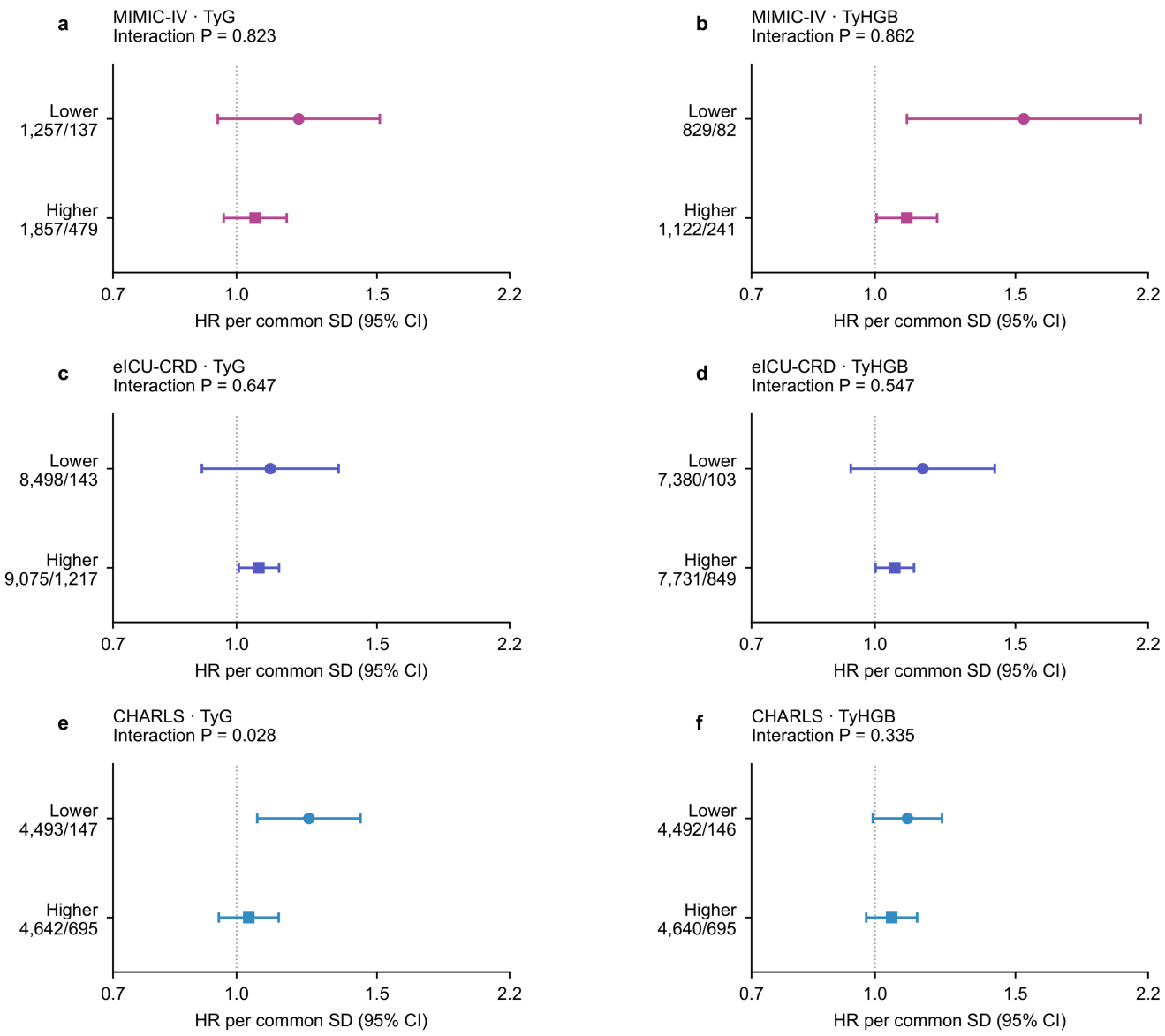


Figure 5

Interaction by severity or age for TyG and TyHGB. Hazard ratios per SD with 95% confidence intervals in lower and higher strata, grouped by cohort: SOFA in MIMIC-IV, APACHE IV in eICU-CRD and age in CHARLS. The same cohort- and index-specific full-sample SD was retained across strata. Cohort colours identify MIMIC-IV, eICU-CRD and CHARLS. Labels show *n*/events; interaction *P* values come from full-sample product-term models, not comparisons of subgroup significance.

Detailed methods for the secondary analyses, and the full results of the dose-response, scaling, proportional-hazards, cut-off, incremental prediction, additive interaction, machine-learning, fasting-status, death-timing, common-adjustment and selection-bias analyses, are provided in Additional file 2. Tables S1-S21 are combined in a single workbook [see Additional file 1], with one sheet per table. Figures S1-S7 are provided as Additional files 3-7 and 9-10; Additional file 7 contains three cohort pages.

Complete figure legends for the additional files are provided in Additional file 2. The analysis code and aggregate results are provided as Additional file 8.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1Tables.xlsx
- Additionalfile2Methods.docx
- Additionalfile3.tiff
- Additionalfile4.tiff
- Additionalfile5.tiff
- Additionalfile6.tiff
- Additionalfile7page1.tiff
- Additionalfile7page2.tiff
- Additionalfile7page3.tiff
- Additionalfile8Code.zip
- Additionalfile9.tiff
- Additionalfile10.tiff
- Additionalfile11STROBEchecklist.pdf
- Additionalfile12TRIPODAIchecklist.pdf