

A three-axis burden score and in-hospital mortality in ICU patients with heart failure and atrial fibrillation: limited incremental discrimination and transportability in a retrospective cohort study

Jiaodan Wang¹

Liaoning University of Traditional Chinese Medicine

Rui Yu

yurui1969@163.com

Liaoning University of Traditional Chinese Medicine

Research Article

Keywords: Heart failure, Atrial fibrillation, Prognostic nutritional index, Neutrophil-to-lymphocyte ratio, Lactate clearance, Intensive care, Risk stratification

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11188724/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

ICU patients with heart failure (HF) and atrial fibrillation (AF) have substantial in-hospital mortality. We defined a 0–3 burden score from a low prognostic nutritional index (PNI), a high neutrophil-to-lymphocyte ratio (NLR) and a low dynamic lactate clearance rate (dLCR), and asked whether the count is associated with in-hospital mortality after adjustment and whether it improves discrimination in a second database.

Methods

Cut-points were the medians of observed values in MIMIC-IV and were applied unchanged in both cohorts. Derivation used MIMIC-IV v3.1 (n = 7,562; 1,222 deaths, 16.2%) and external assessment used eICU-CRD 2.0 (n = 2,588; 414 deaths, 16.0%). Missing values were handled by multiple imputation (m = 10). Associations were odds ratios from logistic models. Incremental discrimination was the change in area under the curve relative to a 43-variable reference model.

Results

In the imputed cohorts, in-hospital mortality rose from 10.8% to 24.9% in MIMIC-IV and from 10.2% to 24.0% in eICU-CRD across scores 0 to 3. Among patients in whom all three axes were measured, mortality was higher at every score, and the eICU-CRD gradient was not monotonic. In MIMIC-IV each additional point had an adjusted odds ratio of 1.20 (95% CI 1.08–1.33). The corresponding estimate in eICU-CRD was 1.14 (0.97–1.35). Out-of-fold, the score raised the AUC of a reference model that already contained albumin, the white-cell count and lactate by 0.0028 (0.0011–0.0045), less than the gain from the axes as continuous terms. That increment was not detectable when the models were refitted in eICU-CRD. The reference model kept its rank order after transport (AUC 0.711) but was miscalibrated; two-parameter recalibration, evaluated out of fold, lowered the Brier score from 0.140 to 0.124.

Conclusions

The count is a descriptive stratification tied to the derivation cut-points. The adjusted association is supported in MIMIC-IV and not in eICU-CRD. It is not a validated or transportable prediction tool, and the absolute risks from the reference model are not available from the three measurements alone.

Background

ICU patients with HF and AF are a high-risk group. Each condition is associated with worse outcomes, and their coexistence adds risk through neurohormonal activation, atrial remodelling and pump failure

[1–3]. Atrial fibrillation during a heart-failure admission is associated with higher in-hospital mortality [3, 5] and with the drivers of repeated hospitalization [4]. Generic severity scores – APACHE II [6], SAPS II [7] and SOFA [8] – were built in mixed ICU populations and were not designed around this cardio-metabolic pattern.

Three measurements available from routine blood tests have each been linked to short-term risk.

Nutritional reserve. The prognostic nutritional index, calculated from albumin and the absolute lymphocyte count, was introduced for surgical risk [9, 10]. Low PNI has been associated with mortality in heart-failure cohorts [11, 12, 28]. In HF, a low albumin also reflects congestion, hepatic synthetic impairment and catabolism, so the index is not a pure measure of intake.

Inflammatory activation. The neutrophil-to-lymphocyte ratio is an inexpensive marker of systemic inflammatory stress [14, 15]. Higher NLR has been associated with short-term mortality in acute HF [16], after myocardial infarction [17], and in AF [18, 19]. In stable HFrEF, NLR and PNI have each been related to disease severity and survival [29].

Perfusion recovery. Dynamic lactate clearance, the relative fall in lactate between two samples, has been associated with survival in severe sepsis and septic shock [20, 21] and in cardiogenic shock [22, 23]. Clearance is not the same quantity as the admission lactate, and a repeat sample is ordered selectively.

Most prognostic work in HF with AF has either fitted a multivariable or machine-learning model to this population [24–26], which can discriminate well but is hard to use at the bedside [27], or has studied one biomarker at a time [11, 12, 16–19, 28]. Composite nutritional scores such as CONUT [13] and P-CONUT [30] need a cholesterol measurement, which is often missing from ICU records.

We asked whether a 0–3 count of low PNI, high NLR and low lactate clearance (i) is associated with in-hospital mortality in ICU patients with HF and AF after adjustment for a critical-care reference model, (ii) improves discrimination over that model by a clinically meaningful amount, (iii) does so consistently across prespecified subgroups, and (iv) whether either the association or the increment is present in a second ICU database. The count uses cut-points frozen in the derivation cohort.

Methods

Design and reporting

Retrospective cohort study with derivation in MIMIC-IV v3.1 and external assessment in eICU-CRD 2.0. Reporting follows STROBE. The logistic components follow TRIPOD [31]. TRIPOD + AI [32] is not used to present an artificial-intelligence predictor; the random-forest summary describes one imputed sample. The primary estimand is the association between the three-axis burden score and in-hospital all-cause mortality. Incremental discrimination is a secondary estimand.

Data sources

MIMIC-IV v3.1 [33] contains de-identified critical-care data from Beth Israel Deaconess Medical Center (2008–2022). eICU-CRD 2.0 [34] contains de-identified data from 208 hospitals in the United States (2014–2015). Cohort construction followed the reproducible MIMIC code repository [35].

Laboratory and vital-sign windows were the same in both databases: from 24 h before ICU admission to 48 h after admission. The window does not lengthen with length of stay. For eICU-CRD, NLR was the ratio of neutrophil percentage to lymphocyte percentage. The absolute lymphocyte count used in PNI was white-cell count $\times$ lymphocyte percentage / 100, restricted to $0.1\text{--}20 \times 10^3/\mu\text{L}$. Percentage NLR equals count NLR when both differentials are taken from the same white-cell count. That identity does not make a MIMIC-IV cell count and an eICU-CRD percentage, produced in different laboratories, the same measurement.

Participants

Adults aged ≥ 18 years with a first ICU stay of at least 24 h and HF and AF recorded on the same hospitalization. In MIMIC-IV, HF was ICD-9 428.* or ICD-10 I50., and AF was ICD-9 427.31 or ICD-10 I48.. In eICU-CRD, HF and AF were taken from the hierarchical diagnosis strings. The codes do not separate new from prevalent AF, or reduced from preserved ejection fraction. Patients with unknown hospital outcome were excluded. In MIMIC-IV the age criterion excluded no one after the diagnosis filter (Fig. 1).

Outcome

In-hospital all-cause mortality, taken from the hospital discharge disposition.

Three-axis burden score

- **Nutritional axis.** PNI = $10 \times \text{albumin (g/dL)} + 5 \times \text{absolute lymphocyte count } (\times 10^3/\mu\text{L})$, which is the Onodera form written with the lymphocyte count in thousands per microlitre. One point if PNI was below 38.0.
- **Inflammatory axis.** NLR = neutrophil percentage / lymphocyte percentage. One point if NLR was at or above 9.24 (log NLR 2.22).
- **Perfusion axis.** dLCR = $(\text{lactate at the earliest sample} - \text{lactate at the latest sample}) / \text{earliest lactate} \times 100\%$. Both samples had to fall inside the -24 h to $+48$ h window and had to have different timestamps. A single lactate was treated as missing, not as zero clearance. Negative clearance, a rising lactate, was retained. One point if dLCR was below 12.5%. The earliest and latest samples inside the window are not a fixed interval: the gap depends on how often lactate was ordered, and a patient who dies early contributes a shorter gap.

The three cut-points are the medians of observed, pre-imputation values in MIMIC-IV. They were frozen and applied to eICU-CRD without re-estimation (frozen_cutpoints.json). Within each cohort the score was calculated on every imputed data set with those same cut-points. A second, descriptive gradient used only stays in which all three axes were observed, with no imputation of the axes.

Reference model

The reference model had 43 variables: age and sex; chronic kidney disease, coronary artery disease, diabetes, chronic obstructive pulmonary disease and hypertension; first heart rate, systolic pressure, mean arterial pressure, respiratory rate, SpO₂ and Glasgow Coma Scale; 24 h minimum and maximum of heart rate, systolic pressure, mean arterial pressure, respiratory rate and SpO₂; admission albumin, lactate, creatinine, urea nitrogen, sodium, potassium, bicarbonate, chloride, glucose, haemoglobin, platelets, red-cell distribution width, calcium, bilirubin, white-cell count and international normalized ratio; mechanical ventilation, vasopressor use and 24 h urine output; and a reconstructed SOFA score. The respiratory SOFA component used SpO₂ because FiO₂ was incompletely recorded. Albumin, white-cell count and lactate level sit in the reference model, so PNI, NLR and dLCR are not independent of it. The Δ AUC below is the gain beyond those components. It is not the discrimination of the three measurements, or of the 0–3 score, used alone. A reference model with albumin, white-cell count and lactate removed was not fitted, and the AUC of the score by itself was not estimated. Either comparison would answer a different question and could show a larger AUC without changing this one.

Statistical analysis

Missing values were imputed by chained equations with Bayesian ridge regression [36], $m = 10$, drawing from the posterior. Continuous variables were winsorised at the 0.1st and 99.9th percentiles before imputation. Logistic coefficients were pooled with Rubin's rules [37].

Associations are odds ratios per one point of the score, and for score 1, 2 and 3 versus score 0. The per-point model treats each extra point as the same increment. That linear form was not tested against the categorical contrasts. The age- and sex-adjusted model is reported separately from the model that adds the score to the 43-variable reference model. The age- and sex-adjusted model is not an unadjusted model.

Discrimination was estimated in three ways, which are not interchangeable. First, a held-out 30% test set in MIMIC-IV was used for the variable-block ladder. Second, 5-fold out-of-fold predictions over the full cohort were used for the incremental comparison of the reference model, the reference model plus the score, and the reference model plus the three axes and their interaction. Δ AUC confidence intervals used 1,000 bootstrap resamples of those out-of-fold predictions [38, 39]. Third, MIMIC-IV coefficients were applied to eICU-CRD. For that transport, a Rubin-pooled coefficient comparison and an imputation-averaged predicted-probability comparison are both reported. Models were also refitted inside eICU-CRD. Net reclassification and integrated discrimination improvement compare the three-axis continuous

specification (A5) with the reference model on the MIMIC-IV test set, at risk cuts of 10% and 20%. They are not estimates for the 0–3 score.

External calibration used the calibration slope and intercept and deciles of predicted risk. Recalibration updated only the intercept and slope of the transported linear predictor. Performance of the recalibrated probabilities was computed from 5-fold cross-fitting, so the slope and intercept are not the in-sample values of a model scored on the same rows used to fit them.

Subgroups were prespecified: age < 65 versus ≥ 65 years, sex, mechanical ventilation, vasopressor use, chronic kidney disease, and SOFA above versus below the cohort median. Twelve strata result. These comparisons are exploratory. Intervals were not adjusted for multiplicity.

Analyses used Python 3 with scikit-learn, statsmodels and SciPy.

Results

Cohorts

MIMIC-IV contributed 7,562 patients (1,222 deaths, 16.2%) and eICU-CRD 2,588 patients (414 deaths, 16.0%) (Fig. 1). In MIMIC-IV, 94,458 ICU stays were linked to 37,517 hospitalizations carrying both HF and AF codes; 13,790 of the ICU stays fell in those hospitalizations; 9,001 were the patient's first stay; 1,439 stays shorter than 24 h were then removed. Counts for the steps between the eICU-CRD source file (200,859 ICU admissions) and the analytic cohort were not retained, and they cannot be rebuilt from the stored export.

Within the analysis window, PNI was observed in 28.8% of MIMIC-IV stays and 61.3% of eICU-CRD stays, NLR in 52.4% and 76.4%, and dLCR in 60.5% and 23.8%. All three axes were observed together in 1,580 MIMIC-IV stays (20.9%) and 448 eICU-CRD stays (17.3%). Mortality in those stays was 25.1% and 24.8%, against 16.2% and 16.0% in the full cohorts. In MIMIC-IV, within the same window, the three axes were all observed in 33.31% of stays that ended in death and in 18.85% of stays in which the patient left hospital alive. Observation was also more common during longer stays. The complete-measurement subset is therefore a sicker, more closely observed group.

Burden score and mortality

Figure 2 and Table 1 separate the imputed full cohort from stays with all three axes measured.

Table 1

In-hospital mortality by burden score Imputed columns are the mean risk across $m = 10$ imputations. The score in those columns uses MIMIC-IV cut-points applied to imputed axis values. Measured columns are stays with PNI, NLR and dLCR all observed; intervals are Wilson 95% confidence intervals. Mean imputed counts need not sum to the cohort size.

Score	MIMIC-IV imputed, n / risk	MIMIC-IV measured, n / deaths / risk (95% CI)	eICU-CRD imputed, n / risk	eICU-CRD measured, n / deaths / risk (95% CI)
0	982 / 10.8%	252 / 33 / 13.1% (9.5–17.8)	480 / 10.2%	86 / 14 / 16.3% (10.0–25.5)
1	2,578 / 12.7%	528 / 100 / 18.9% (15.8–22.5)	820 / 13.1%	125 / 31 / 24.8% (18.1–33.1)
2	2,763 / 17.4%	548 / 152 / 27.7% (24.2–31.6)	888 / 18.2%	169 / 39 / 23.1% (17.4–30.0)
3	1,240 / 24.9%	252 / 111 / 44.1% (38.1–50.2)	399 / 24.0%	68 / 27 / 39.7% (28.9–51.6)

In the imputed cohorts, mortality rose with the score in both databases: MIMIC-IV 10.8%, 12.7%, 17.4% and 24.9%; eICU-CRD 10.2%, 13.1%, 18.2% and 24.0%. The absolute difference from score 0 to score 3 was 14.1 and 13.8 percentage points.

Where all three axes were measured, MIMIC-IV mortality was 13.1%, 18.9%, 27.7% and 44.1%. eICU-CRD mortality was 16.3%, 24.8%, 23.1% and 39.7%. The eICU-CRD series is not monotonic: score 1 sits above score 2, and the Wilson intervals overlap. Score 3 in that subset contains 68 stays and 27 deaths.

Age- and sex-adjusted odds ratios per point were 1.41 (1.30–1.53) in MIMIC-IV and 1.40 (1.23–1.60) in eICU-CRD. After the reference model, the MIMIC-IV odds ratio was 1.20 (1.08–1.33; $P < 0.001$). The eICU-CRD odds ratio was 1.14 (0.97–1.35; $P = 0.120$). Against score 0 in the fully adjusted MIMIC-IV model, score 1 was 1.10 (0.79–1.53; $P = 0.579$), score 2 was 1.17 (0.83–1.64; $P = 0.363$), and score 3 was 1.74 (1.20–2.54; $P = 0.004$). The eICU-CRD contrast of score 3 versus 0 was 1.56 (0.90–2.72; $P = 0.116$) (Table 2).

Table 2

Association of the burden score with in-hospital mortality Odds ratios are Rubin-pooled across $m = 10$ imputations. Adjusted models add the score to the 43-variable reference model. Age- and sex-adjusted models are not unadjusted.

Cohort	Contrast	OR	95% CI	P
MIMIC-IV	Per point, age and sex	1.41	1.30–1.53	< 0.001
MIMIC-IV	Per point, adjusted	1.20	1.08–1.33	< 0.001
MIMIC-IV	Score 1 vs 0, adjusted	1.10	0.79–1.53	0.579
MIMIC-IV	Score 2 vs 0, adjusted	1.17	0.83–1.64	0.363
MIMIC-IV	Score 3 vs 0, adjusted	1.74	1.20–2.54	0.004
eICU-CRD	Per point, age and sex	1.40	1.23–1.60	< 0.001
eICU-CRD	Per point, adjusted	1.14	0.97–1.35	0.120
eICU-CRD	Score 3 vs 0, adjusted	1.56	0.90–2.72	0.116

Table 3 is a standardized risk from the reference model: every patient is assigned each score in turn while the other covariates stay as observed. It is not the mortality a clinician can read from the three laboratory values alone.

Table 3

Standardized in-hospital mortality by assigned score Each patient was forced to each score while other covariates were left unchanged. Predictions come from the reference model plus the score and are averaged across imputations. The parenthetical values are the standard deviation of that mean across the 10 imputations, not confidence intervals. These risks cannot be read off the three axis values without the reference model.

Score	MIMIC-IV standardized risk	eICU-CRD standardized risk
0	12.9% (0.46)	13.6% (0.72)
1	14.8% (0.21)	15.0% (0.31)
2	16.8% (0.10)	16.5% (0.17)
3	19.0% (0.48)	18.2% (0.71)

Discrimination

On the held-out 30% of MIMIC-IV, logistic AUCs were 0.588 for age and sex, 0.608 after comorbidities, 0.694 after vital signs, 0.762 after laboratory values, 0.772 after organ support, and 0.773 for the full reference model. The largest block increment was vital signs (+ 0.086), then laboratory values (+ 0.068).

Out-of-fold over the full MIMIC-IV cohort, the reference AUC was 0.772. Adding the continuous axes and their interaction gave 0.776, a Δ AUC of + 0.0046 (0.0021 to 0.0072). Adding the burden score gave 0.774, a Δ AUC of + 0.0028 (0.0011 to 0.0045). The score increment was smaller than the continuous-axis increment (Fig. 3). The in-sample AUC of the reference model on all 7,562 MIMIC-IV stays was 0.783; that figure is not cross-validated.

On the MIMIC-IV test set, categorical net reclassification for the continuous three-axis specification versus the reference model was - 0.0005 (- 0.031 to 0.031). Integrated discrimination improvement was + 0.0094 (0.0039 to 0.0157). The reclassification interval includes zero.

Subgroups

The subgroup comparisons are exploratory. Point estimates of the score's Δ AUC were positive in all 12 MIMIC-IV subgroups, from + 0.0025 to + 0.0041 (Fig. 5). The bootstrap interval excluded zero for age $\geq$ 65 years, male sex, mechanical ventilation, vasopressor use, no vasopressor, no chronic kidney disease, and SOFA at or above the median. It included zero for age < 65 years, female sex, no mechanical ventilation, chronic kidney disease, and SOFA below the median. In eICU-CRD every one of the 12 intervals included zero. Point estimates there ranged from - 0.0057 to + 0.0044.

External transport and recalibration

Applied to eICU-CRD, the MIMIC-IV reference model had AUC 0.711 and Brier score 0.140, with calibration slope 0.70 and intercept - 0.95. The upper decile illustrates the miscalibration: mean predicted mortality was 67.6% and observed mortality was 38.6% (Fig. 4). After two-parameter recalibration evaluated out of fold, AUC was 0.710, Brier score 0.124, slope 0.99 and intercept - 0.01. Refitting the reference model inside eICU-CRD gave AUC 0.729.

Transport of the continuous-axis model, using Rubin-pooled coefficients, increased AUC by 0.0053 (0.0015 to 0.0090; $P = 0.006$) relative to the transported reference model. Averaging predictions across imputations instead of pooling coefficients gave an increment of 0.0057. These increments belong to the continuous specification, not to the 0–3 score.

When both models were refitted inside eICU-CRD, out-of-fold Δ AUC was + 0.0009 for the score (- 0.0024 to 0.0042) and + 0.0009 for the continuous axes (- 0.0025 to 0.0038).

The three axes separately

Per 1 SD, in age- and sex-adjusted MIMIC-IV models, PNI, log NLR and dLCR were all associated with mortality in the expected direction: odds ratios 0.78 (0.71–0.87), 1.30 (1.16–1.45) and 0.82 (0.77–0.87). After the reference model, PNI was 0.95 (0.78–1.16; $P = 0.624$) and log NLR was 1.09 (0.95–1.25; $P = 0.233$). dLCR remained associated: 0.79 (0.73–0.85; $P < 0.001$).

In eICU-CRD the age- and sex-adjusted odds ratios were 0.80 (0.69–0.93) for PNI, 1.23 (1.07–1.41) for log NLR, and 0.89 (0.79–1.02; $P = 0.084$) for dLCR. None of the three remained significant after the

reference model (Additional file 1: Table S2).

Discussion

What the two cohorts support

In imputed data, in-hospital mortality rose with the three-axis count in both ICU databases, by about 14 percentage points from score 0 to score 3 (Fig. 2). That description does not require the reference model. It also does not survive unchanged once the score is restricted to stays with all three measurements: those patients had higher mortality at every score, and in eICU-CRD the ordering of scores 1 and 2 reversed.

The adjusted association is a MIMIC-IV result. One extra point corresponded to an odds ratio of 1.20 after 43 covariates. Scores 1 and 2 were not distinguishable from score 0 in that model; score 3 was. In eICU-CRD the adjusted per-point interval included 1. The twelve subgroup increments are exploratory: intervals that exclude zero were limited to seven MIMIC-IV strata and to none of the eICU-CRD strata.

The reference model's discrimination transported, with an AUC of 0.711 against an out-of-fold MIMIC-IV AUC of 0.772, and its calibration did not. Two-parameter recalibration, scored out of fold, reduced the Brier score from 0.140 to 0.124 and brought the slope and intercept close to the ideal values. That repair applies to the reference model. The 0–3 score did not add detectable discrimination when the models were refitted in eICU-CRD. The continuous-axis specification did show a small AUC gain under coefficient transport (+ 0.005). That gain is not a property of the bedside count.

The Δ AUC of 0.0028 is not the accuracy of the score used by itself. Albumin, the white-cell count and the lactate level are already among the 43 covariates, so the increment is only the information left after those components. Removing them, or reporting an AUC for the 0–3 count alone, would describe a weaker comparator and could produce a larger number. That number would not mean the count improves on routine laboratories. Calibration was estimated for the transported reference model, not for a model that contains only the score.

What the axes contribute

Entering PNI, NLR and dLCR as continuous terms improved MIMIC-IV discrimination by more than dichotomising them and summing the points. The data do not support the claim that the 0–3 score isolates a shared latent state better than the continuous measurements. A latent-factor analysis was not done.

The axis that remained after full adjustment in MIMIC-IV was lactate clearance, not NLR. PNI and NLR largely overlapped information already in albumin, the white-cell count and the rest of the reference model. dLCR uses two lactate values, whereas the reference model contains the lactate level, so a residual association is plausible. It was not confirmed after adjustment in eICU-CRD, where dLCR was

observed in only 23.8% of stays. Age- and sex-adjusted associations in both cohorts pointed the same way for PNI and NLR, which is a weaker statement than independent prognostic value.

CONUT and P-CONUT need cholesterol [13, 30]. The burden score does not. In exchange it needs a second lactate inside a fixed window, which was absent for most eICU-CRD stays and for 39.5% of MIMIC-IV stays.

How the score can and cannot be used

A site that accepts the MIMIC-IV cut-points (PNI 38, NLR 9.24, dLCR 12.5%) can compute the 0–3 count from those three quantities. The count is a flag, not a predicted probability. Table 1’s imputed risks describe these two databases. Table 1’s complete-measurement risks describe a sicker subset. Table 3’s standardized risks require the 43-variable model and the imputation machinery; they are not a bedside card. Nothing in this study tests whether acting on a high score — more monitoring, nutrition support, or a change in resuscitation — changes outcome.

Because the cut-points were frozen in MIMIC-IV, applying local medians would define a different score. We did not re-estimate the cut-points in eICU-CRD, including as a sensitivity analysis. A locally cut score might fit that database more closely. It would no longer be the score locked in MIMIC-IV, and it is not reported here.

Limitations

The design is retrospective. Measurement of a repeat lactate is itself a clinical decision, so dLCR mixes physiology with indication. Multiple imputation assumes that missingness is conditional on observed variables; if laboratories are ordered because a patient looks sicker, and that judgement is not in the covariates, the imputed gradient can be biased. The reconstructed SOFA is not the original SOFA. Administrative codes miss some HF and AF and cannot type either condition. Ejection fraction, AF pattern and medication exposure were not in the analysis. The external reference-model refit has about 10 events per variable (414 events, 43 covariates), so its coefficients are imprecise and are not interpreted further. The eICU-CRD flow table does not retain the counts removed between the source file and the analytic cohort, and those counts cannot be rebuilt from the export. Incremental AUC values of 0.003 to 0.005 will not reclassify most patients, and the categorical reclassification index was null. The twelve subgroups were not adjusted for multiplicity. We did not raise the number of imputations above 10, add a missing-indicator term, or calibrate a score-only model. Those checks need the patient-level records, which are not among the stored result tables. Table 3 remains a standardized risk from the full model. It cannot be read from the three laboratory values, and nothing in the study tests a change in monitoring or nutrition.

Conclusions

A three-axis count of low PNI, high NLR and low lactate clearance, using cut-points frozen in MIMIC-IV, tracks in-hospital mortality in imputed data from two ICU cohorts of patients with HF and AF. After a 43-

variable reference model, the association is supported in MIMIC-IV and not in eICU-CRD. The count adds a very small amount of discrimination in the derivation data, less than the continuous axes, and that increment is not detectable on refitting in eICU-CRD. Discrimination of the reference model transports and its calibration can be repaired with an intercept and a slope. The count is a stratification description tied to the derivation cut-points, not a validated prediction instrument.

Abbreviations

AF, atrial fibrillation; AUC, area under the receiver operating characteristic curve; dLCR, dynamic lactate clearance rate; HF, heart failure; ICD, International Classification of Diseases; NLR, neutrophil-to-lymphocyte ratio; OR, odds ratio; PNI, prognostic nutritional index; SOFA, Sequential Organ Failure Assessment.

Declarations

Ethics approval and consent to participate

MIMIC-IV and eICU-CRD are de-identified critical-care databases. Both were used under PhysioNet data-use agreements. The present analysis of these secondary data did not require additional institutional review or individual consent.

Consent for publication

Not applicable.

Clinical trial number

Clinical trial number: not applicable.

Availability of data and materials

The data that support the findings of this study are available from PhysioNet (MIMIC-IV and the eICU Collaborative Research Database) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of PhysioNet. The aggregate tables underlying the estimates in this article are included in Additional file 1. Analysis code is available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding was received for this study.

Authors' contributions

Jiaodan Wang collected the data from the MIMIC database, carried out statistical analysis, produced figures and drafted the manuscript. Rui Yu conceived the study and critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements

We thank the MIT Laboratory for Computational Physiology and the contributors to the eICU Collaborative Research Database.

References

1. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure. *Circulation*. 2022;145:e895–1032. 10.1161/CIR.0000000000001063.
2. Wang TJ, Larson MG, Levy D, et al. Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: the Framingham Heart Study. *Circulation*. 2003;107:2920–5. 10.1161/01.CIR.0000072767.89944.6E.
3. Mountantonakis SE, Grau-Sepulveda M, Bhatt DL, et al. Atrial fibrillation is associated with higher in-hospital mortality in patients hospitalized with heart failure and reduced ejection fraction. *J Am Heart Assoc*. 2016;5:e003580. 10.1161/JAHA.116.003580.
4. Steinberg BA, Kim S, Fonarow GC, et al. Drivers of hospitalization in patients with heart failure and atrial fibrillation: an EVEREST analysis. *JACC Heart Fail*. 2014;2:407–14. 10.1016/j.jchf.2014.02.013.
5. Yan M, Liu H, Xu Q, et al. Development and validation of a prediction model for in-hospital death in patients with heart failure and atrial fibrillation. *BMC Cardiovasc Disord*. 2023;23:142. 10.1186/s12872-023-03166-0.
6. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13:818–29. 10.1097/00003246-198510000-00009.
7. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA*. 1993;270:2957–63. 10.1001/jama.1993.27020290036033.
8. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med*. 1996;22:707–10. 10.1007/BF01709751.

9. Buzby GP, Mullen JL, Matthews DC, Hobbs CL, Rosato EF. Prognostic nutritional index in gastrointestinal surgery. *Am J Surg.* 1980;139:160–7. 10.1016/0002-9610(80)90246-9.
10. Onodera T, Goseki N, Kosaki G. Prognostic nutritional index in gastrointestinal surgery of malnourished cancer patients. *Nihon Geka Gakkai Zasshi.* 1984;85:1001–5.
11. Cheng YL, Sung SH, Cheng HM, et al. Prognostic nutritional index and the risk of mortality in patients with acute heart failure. *J Am Heart Assoc.* 2017;6:e004876. 10.1161/JAHA.116.004876.
12. Hu Y, Cao Q, Wang H, et al. Prognostic nutritional index as a predictor of mortality in heart failure: a meta-analysis. *Nutrition.* 2023;114:112098. 10.1016/j.nut.2023.112098.
13. de Ulíbarri JI, González-Madroño A, González P, et al. CONUT: a tool for controlling nutritional status. *Nutr Hosp.* 2005;20:38–45.
14. Zahorec R. Ratio of neutrophil to lymphocyte counts – rapid and simple parameter of systemic inflammation and stress in critically ill. *Bratisl Lek Listy.* 2001;102:5–14.
15. Diakos CI, Charles KA, McMillan DC, Clarke SJ. Cancer-related inflammation and treatment effectiveness. *Lancet Oncol.* 2014;15:e493–503. 10.1016/S1470-2045(14)70263-3.
16. Uthamalingam S, Patvardhan EA, Subramanian S, et al. Utility of the neutrophil-to-lymphocyte ratio to predict short-term mortality in patients with acute heart failure. *Am J Cardiol.* 2011;107:433–8. 10.1016/j.amjcard.2010.09.039.
17. Tamhane UU, Aneja S, Montgomery D, et al. Association between admission neutrophil to lymphocyte ratio and in-hospital mortality after acute myocardial infarction. *Am J Cardiol.* 2008;102:653–7. 10.1016/j.amjcard.2008.05.006.
18. Durmus E, Kivrak T, Gerin F, et al. Neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio are predictors of heart failure, all-cause and cardiovascular mortality in patients with atrial fibrillation. *Int J Cardiol.* 2023;389:129968. 10.1016/j.ijcard.2023.129968.
19. Angkananard T, Inthano T, Chanchai R, et al. Neutrophil-to-lymphocyte ratio and cardiovascular outcomes in patients with atrial fibrillation: a systematic review and meta-analysis. *Front Cardiovasc Med.* 2024;11:1461080. 10.3389/fcvm.2024.1461080.
20. Nguyen HB, Rivers EP, Knoblich BP, et al. Early lactate clearance is associated with improved outcome in severe sepsis and septic shock. *Crit Care Med.* 2004;32:1637–42. 10.1097/01.CCM.0000132904.35713.A7.
21. Jones AE, Shapiro NI, Trzeciak S, et al. Lactate clearance vs central venous oxygen saturation as goals of early sepsis therapy: a randomized clinical trial. *JAMA.* 2010;303:739–46. 10.1001/jama.2010.158.
22. Bakker J, De Backer D, Hernandez G. Lactate clearance in cardiogenic shock: a target for intervention, a marker of mortality, or both? *J Crit Care.* 2019;50:151–7. 10.1016/j.jcrc.2018.11.014.
23. Park IH, Yang JH, Jang WJ, et al. Clinical significance of lactate clearance in patients with cardiogenic shock: results from the RESCUE registry. *J Intensive Care.* 2021;9:63. 10.1186/s40560-021-00571-7.

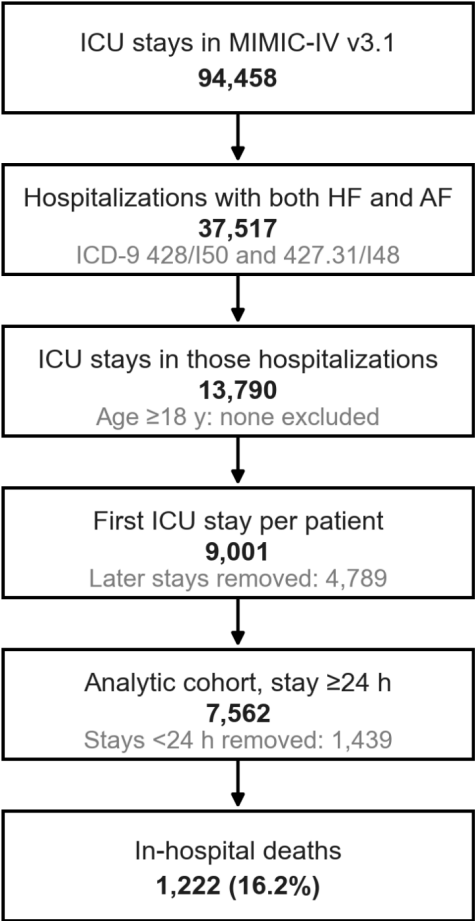
24. Chen P, Sun J, Chu Y, et al. Predicting in-hospital mortality in patients with heart failure combined with atrial fibrillation using stacking ensemble model: an analysis of the MIMIC-IV database. *Front Cardiovasc Med.* 2023;10:1232494. 10.3389/fcvm.2023.1232494.
25. Chen Z, Li T, Guo S, et al. Machine learning-based in-hospital mortality risk prediction tool for intensive care unit patients with heart failure. *Front Cardiovasc Med.* 2023;10:1118187. 10.3389/fcvm.2023.1118187.
26. Cai D, Chen Q, Mu X, et al. Development and validation of a novel combinatorial nomogram model to predict in-hospital deaths in heart failure patients. *BMC Cardiovasc Disord.* 2024;24:84. 10.1186/s12872-024-03718-0.
27. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data. *JAMA Intern Med.* 2016;176:828–30. 10.1001/jamainternmed.2016.1145.
28. Zhang H, Tao M, Sun L, et al. Association of the prognostic nutritional index with mortality in patients with heart failure: a systematic review and meta-analysis. *Front Cardiovasc Med.* 2024;11:1392573. 10.3389/fcvm.2024.1392573.
29. Arfsten H, Cho A, Prausmüller S, et al. Inflammation-based scores as a common tool for prognostic assessment in heart failure or cancer. *Front Cardiovasc Med.* 2021;8:725903. 10.3389/fcvm.2021.725903.
30. Li J, et al. Enhanced prognostic value of a composite nutritional-inflammatory index (P-CONUT) for predicting mortality risk in patients initiating peritoneal dialysis. *J Ren Nutr.* 2022;32:442–52. 10.1053/j.jrn.2021.06.012.
31. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162:55–63. 10.7326/M14-0697.
32. Collins GS, Moons KGM, Debray TPA, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or artificial intelligence methods. *BMJ.* 2024;385:e078378. 10.1136/bmj-2023-078378.
33. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data.* 2023;10:1. 10.1038/s41597-022-01899-x.
34. Pollard TJ, Johnson AEW, Raffa JD, et al. 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.
35. Johnson AEW, Stone DJ, Celi LA, Pollard TJ. The MIMIC Code Repository: enabling reproducibility in critical care research. *J Am Med Inf Assoc.* 2018;25:32–9. 10.1093/jamia/ocx098.
36. van Buuren S. *Flexible Imputation of Missing Data*. 2nd ed. Boca Raton: Chapman and Hall/CRC; 2018. 10.1201/9780429492259.
37. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley; 1987. 10.1002/9780470316696.

38. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman and Hall; 1993.

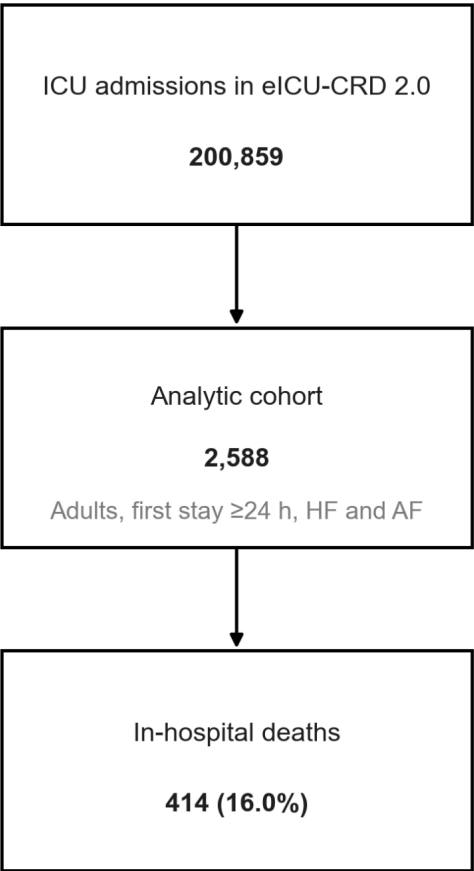
39. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. 2nd ed. New York: Springer; 2019. 10.1007/978-3-030-16399-0.

Figures

a MIMIC-IV



b eICU-CRD



Steps between the source file and this cohort were not stored.

Figure 1

Cohort flow in MIMIC-IV and eICU-CRD.

a MIMIC-IV v3.1. Hospitalizations with both HF and AF were identified with ICD-9 428.* or ICD-10 I50., and ICD-9 427.31 or ICD-10 I48.. Age ≥18 years excluded no one after that filter. **b** eICU-CRD 2.0. The

exported flow retains the source-file size and the analytic cohort only. The last box in each panel is in-hospital deaths.

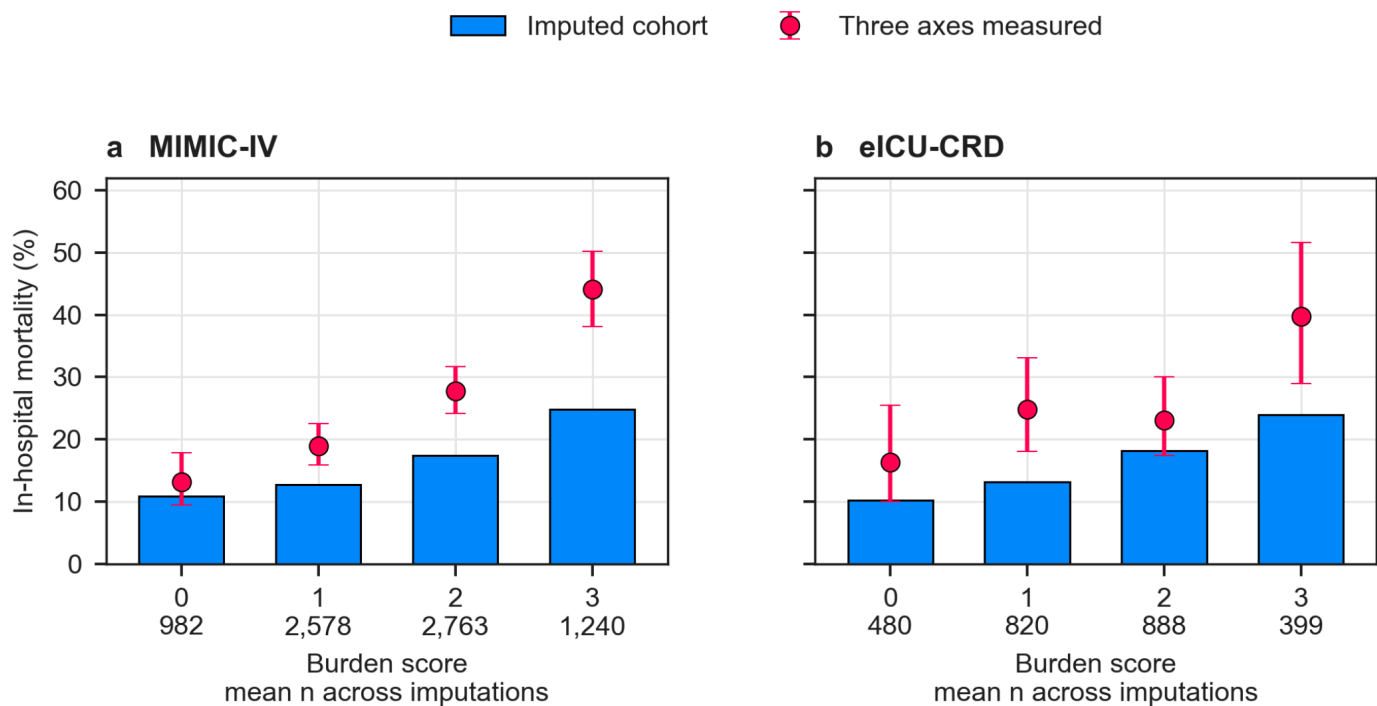


Figure 2

In-hospital mortality by three-axis burden score.

aMIMIC-IV. **b** eICU-CRD. Bars are the mean observed mortality in the imputed cohort. Points are stays with PNI, NLR and dLCR all measured, with Wilson 95% confidence intervals. Counts under the axis are mean imputed sample sizes. The eICU-CRD measured series is not monotonic.

● Interval excludes 0 ◇ Interval includes 0

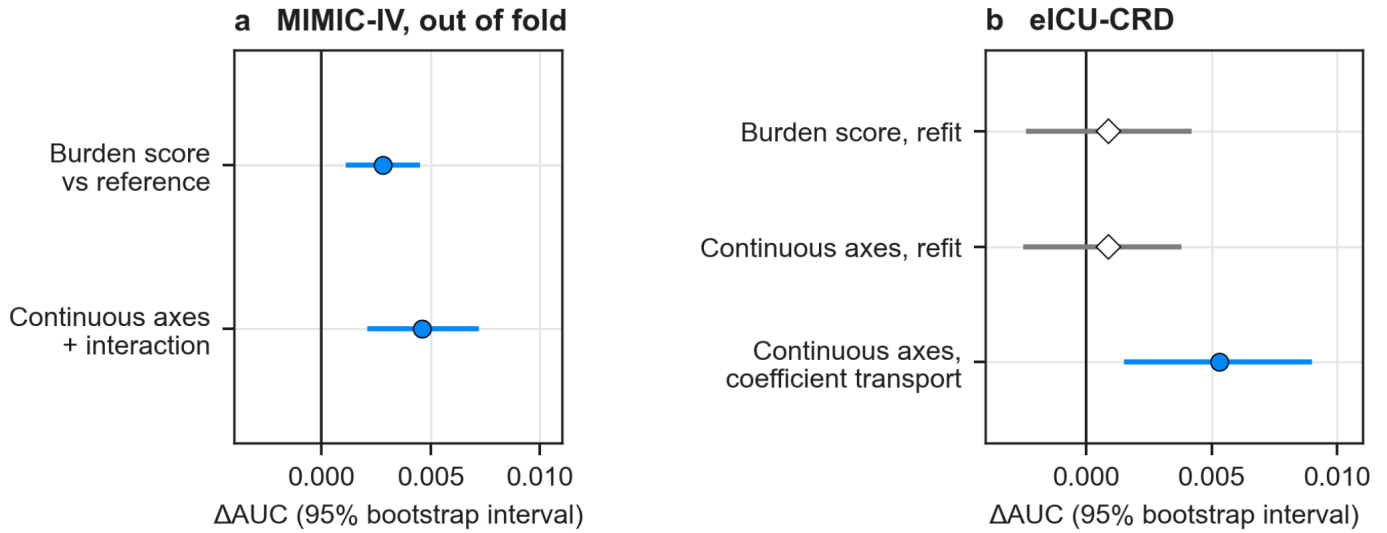


Figure 3

Incremental discrimination beyond the reference model.

Δ AUC versus the 43-variable reference model, with 95% bootstrap intervals. **a** MIMIC-IV, 5-fold out-of-fold predictions, 1,000 resamples. **b** eICU-CRD. Refit rows are out-of-fold predictions inside eICU-CRD. The transport row uses Rubin-pooled MIMIC-IV coefficients, with a 2,000-resample interval. Filled circles: the interval excludes zero. Open diamonds: the interval includes zero.

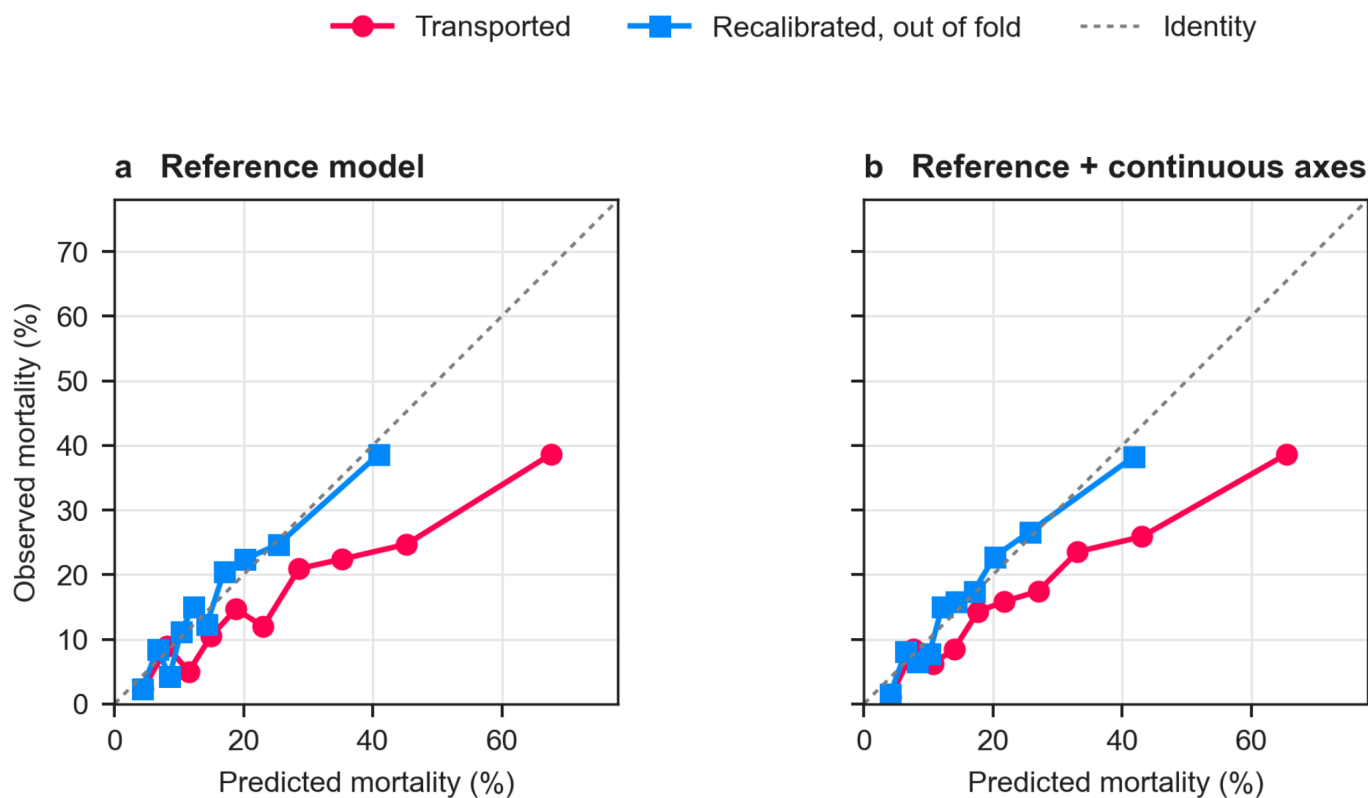


Figure 4

Calibration of transported models in eICU-CRD.

Deciles of predicted risk. **a**Reference model. **b** Reference model plus the continuous axes and their interaction. The dashed line is identity. Transported probabilities overstate risk in the upper deciles (predicted mortality 67.6% and 65.4% in the top decile). Recalibrated probabilities are out of fold.

● Interval excludes 0 ◇ Interval includes 0

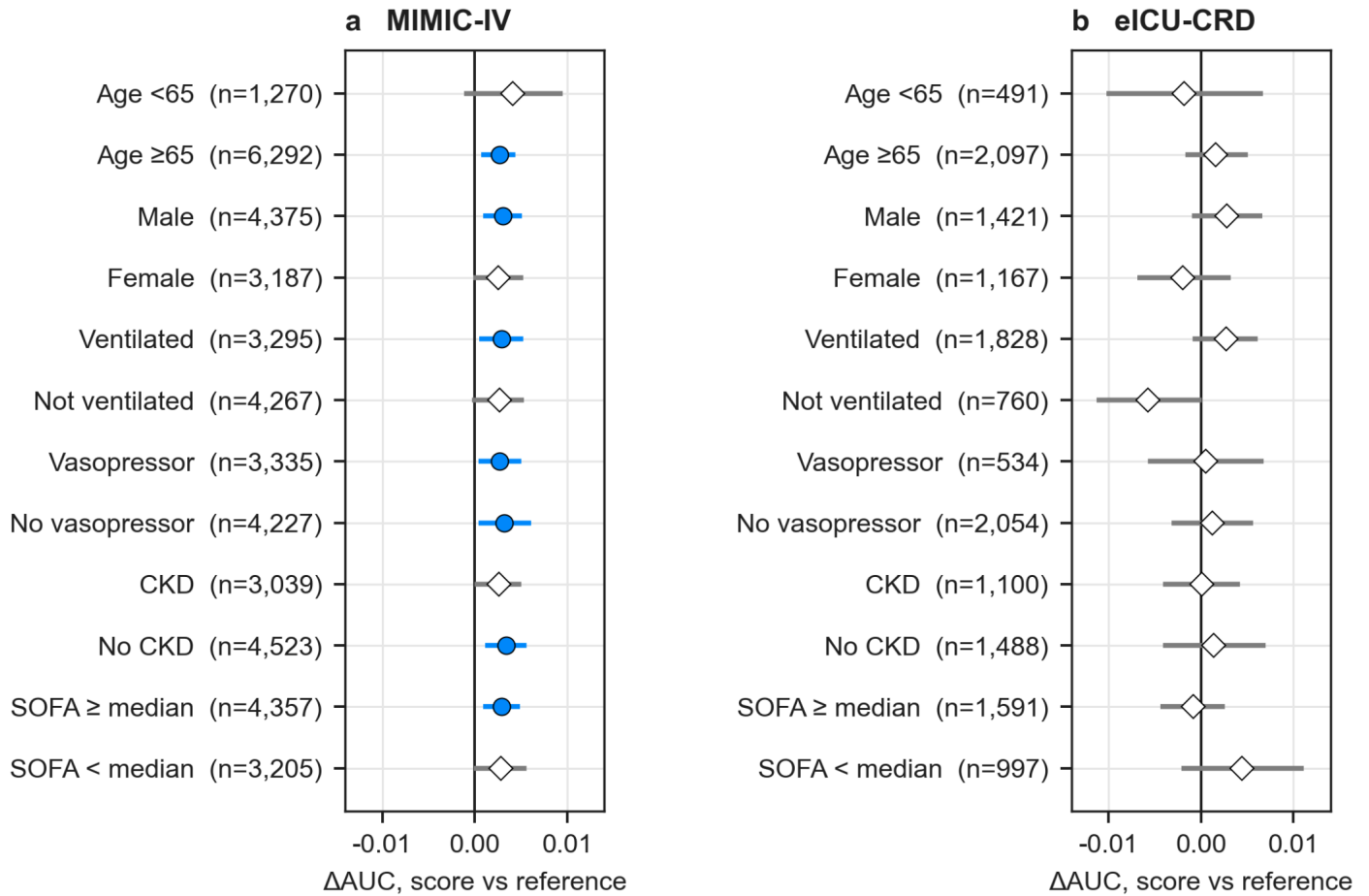


Figure 5

Subgroup change in AUC for the burden score.

Intervals are 95% bootstrap intervals and are not multiplicity-adjusted. **a** MIMIC-IV. **b** eICU-CRD. Filled circles: the interval excludes zero. Open diamonds: the interval includes zero.

Supplementary Files

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

- [Supplementarymaterial.docx](#)