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.

First-day heart rate improves post-landmark mortality risk stratification in critically ill adults with type 2 diabetes and acute myocardial infarction: development in MIMIC-IV and multicentre validation in eICU

hongwei tan
Affiliated Hospital of Southwest Medical University
Zheng Liu
Affiliated Hospital of Southwest Medical University
Zhuo Chen
Affiliated Hospital of Southwest Medical University
Haozhong Huang
Affiliated Hospital of Southwest Medical University
Tan Yang
Affiliated Hospital of Southwest Medical University
Lei Li
Affiliated Hospital of Southwest Medical University
Yongmei Nie
Affiliated Hospital of Southwest Medical University
Fengxu Yu
yuluchuan@163.com

Affiliated Hospital of Southwest Medical University

Research Article

Keywords: acute myocardial infarction, type 2 diabetes, heart rate, landmark analysis, mortality, MIMIC-IV, eICU, external validation

Posted Date: September 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10969848/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

Risk assessment in acute myocardial infarction (AMI) and type 2 diabetes should respect the time at which clinical information becomes available. We evaluated whether mean heart rate during the first 24 hours improves prediction of subsequent in-hospital mortality beyond a compact cross-database clinical model.

Methods

We performed a retrospective 24-hour landmark study using MIMIC-IV for development and the multicentre eICU Collaborative Research Database for external validation. Adults with AMI and type 2 diabetes who remained in intensive care at 24 hours were eligible. The reference logistic model included age, sex, creatinine, and mean oxygen saturation; the augmented model added mean heart rate. Preprocessing and coefficients were frozen in MIMIC-IV. Internal validity was assessed with 500 bootstrap resamples and external uncertainty with 1000 hospital-cluster bootstrap resamples.

Results

The landmark cohorts comprised 2193 MIMIC-IV patients (336 deaths; 15.3%) and 287 eICU patients from 65 hospitals (35 deaths; 12.2%). Bootstrap-corrected development AUROC was 0.660 for the reference model and 0.690 for the heart-rate-augmented model. In eICU, AUROC increased from 0.691 (95% CI 0.623–0.798) to 0.742 (95% CI 0.671–0.849); the paired increase was 0.051 (95% CI 0.001-0.100). The augmented model had an AUPRC of 0.284, Brier score of 0.099, calibration intercept of -0.303, and calibration slope of 1.187. Adding systolic blood pressure change did not improve performance in the complete-window sensitivity cohort. On a common eICU subset, APACHE IVa achieved an AUROC of 0.809 versus 0.739 for the augmented model, with an imprecise difference.

Conclusions

A first-day mean heart rate provided a modest, externally validated improvement in post-landmark mortality risk stratification using routinely recorded variables. The model did not outperform APACHE IVa and requires prospective validation before clinical use.

Research Insights

Element	Highlight
What is currently known?	Diabetes increases cardiovascular risk after AMI.
	First-day physiology may add information beyond admission variables.
Key research question	Does 24-h mean heart rate improve post-landmark mortality prediction?
What is new?	External AUROC increased by 0.051 after adding mean heart rate.
	Delta systolic pressure did not provide stable incremental value.
	The compact model did not outperform APACHE IVa.
Clinical influence	Use as a research-grade reassessment model, not a deployment-ready score.

Background

Acute myocardial infarction (AMI) remains a time-sensitive cardiovascular emergency, while type 2 diabetes adds a high-risk substrate of diffuse atherosclerosis, renal dysfunction, autonomic disturbance, and metabolic vulnerability [1, 2]. Diabetic patients may have less typical symptoms and worse short-term outcomes, but symptom fields in retrospective electronic health records (EHRs) are often too incomplete for reliable phenotype adjudication [3]. Risk stratification should therefore rest on reproducible clinical data rather than presumed absence of symptoms.

Established acute coronary and critical-illness scores provide valuable prognostic context [4–7]. Their variables, however, may not all be available or identically encoded across ICU databases. Repeated vital signs are among the most transportable EHR measurements and may represent evolving shock, autonomic tone, treatment response, arrhythmia, or mechanical complications. Recent multi-database work also underscores that the temporal dimension of physiologic exposure can be clinically informative [8].

A central design problem is temporal alignment. A model that uses measurements collected over 24 hours cannot validly predict deaths occurring before those measurements are complete. We therefore used a 24-hour landmark and evaluated subsequent in-hospital mortality. Our primary objective was to determine whether first-day mean heart rate adds externally validated prognostic information to a compact clinical model in adults with AMI and type 2 diabetes. Secondary objectives were to assess calibration and net benefit, test SBP change and liver variables separately, and benchmark performance against APACHE IVa.

Methods

Study design and data sources

This retrospective prognostic model study used MIMIC-IV for development and the eICU Collaborative Research Database for multicentre external validation. MIMIC-IV contains deidentified hospital and ICU data from Beth Israel Deaconess Medical Center; eICU contains data from participating US hospitals in the Philips eICU programme [9, 10]. Both databases are available through PhysioNet to credentialed researchers under data-use agreements [11]. Reporting was structured around TRIPOD + AI and risk-of-bias principles [12, 13].

The study question was defined as a 24-hour landmark prediction task. Information recorded during the first 24 hours after ICU admission was used to predict in-hospital death occurring after the landmark. This design prevents predictors recorded after early deaths from being used to predict those deaths and positions the model as a first-day reassessment rather than an admission score.

Participants and phenotype mapping

We selected adults aged 18 years or older with an index ICU encounter containing structured evidence of AMI and type 2 diabetes. In MIMIC-IV, AMI was identified with ICD-9-CM 410.x1 or ICD-10-CM I21-I22 codes and type 2 diabetes with ICD-9-CM 250.x0/250.x2 or ICD-10-CM E11 codes. In eICU, diagnosis strings were mapped to acute myocardial infarction and type II diabetes concepts. We retained the first ICU stay per hospitalisation in MIMIC-IV and the first ICU unit stay per hospitalisation in eICU.

Patients were required to remain in the ICU at 24 hours; eICU patients were additionally required to remain in hospital at that time. Heart rate and oxygen saturation measurements during hours 0–24 were required. A planned absence-of-chest-pain phenotype was not used because audit of the available diagnosis fields showed that non-documentation could not adjudicate symptom absence consistently across databases.

Outcome and predictor window

The outcome was subsequent in-hospital mortality, defined from hospital discharge status among patients alive and under observation at the 24-hour landmark. Candidate predictors were selected before modelling for clinical plausibility, first-day availability, and cross-database reproducibility.

The reference model included age, sex, the first creatinine value, and mean peripheral oxygen saturation (SpO2) during hours 0–24. The augmented model added mean heart rate during the same interval. Age and heart rate were scaled per 10 units; creatinine was log2 transformed. The reference specification was deliberately compact relative to the 336 development events and 35 external events [14].

Preprocessing and model estimation

All preprocessing parameters were estimated in MIMIC-IV and frozen before external validation. Continuous predictors were winsorised at the MIMIC-IV 1st and 99th percentiles. Missing creatinine

values were replaced by the MIMIC-IV median (1.2 mg/dL); age, sex, mean SpO2, and mean heart rate were complete in the final landmark cohort. Logistic regression coefficients were estimated in MIMIC-IV without re-estimation or recalibration in eICU.

For reproducibility, the augmented model linear predictor was: -6.593879 + 0.354741×(age/10) + 0.134148×female + 0.497314×log2(creatinine) − 0.004665×mean SpO2 + 0.284407×(mean heart rate/10). Predicted risk equals exp(linear predictor)/(1 + exp(linear predictor)). Sex was coded male = 0 and female = 1. The submitted code and preprocessing table should be used rather than manual transcription for implementation.

Model performance and uncertainty

Discrimination was quantified by the area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC). Overall prediction error was quantified by the Brier score. Calibration was assessed with calibration-in-the-large (intercept), calibration slope, and grouped calibration plots [15]. Decision-curve analysis evaluated net benefit across threshold probabilities of 0.05–0.30 [16].

Internal validity was assessed with 500 bootstrap resamples in which the complete modelling pipeline was re-estimated and optimism subtracted from apparent performance [17]. External 95% confidence intervals and paired model differences were calculated from 1000 bootstrap resamples of eICU hospitals, retaining within-hospital correlation.

Sensitivity and benchmark analyses

First, an expanded model added log2-transformed alanine aminotransferase, aspartate aminotransferase, bilirubin and alkaline phosphatase plus albumin; development medians were used for imputation because these variables had substantial missingness. Second, systolic blood pressure (SBP) change was defined as mean SBP at 18–24 hours minus mean SBP at 0–6 hours and tested only in patients with both windows. Third, category-based net reclassification improvement (NRI) and integrated discrimination improvement (IDI) were treated as sensitivity measures [18]. Finally, the augmented model was compared with the available APACHE IVa hospital mortality probability on a common eICU subset [19]. No claim of superiority was prespecified.

Continuous variables are reported as median (interquartile range) and categorical variables as number (percentage). Analyses were conducted in R 4.6.1. The analysis date was 29 August 2026.

Results

Study cohorts

MIMIC-IV contained 2547 index ICU stays with AMI and type 2 diabetes; 2196 patients remained in the ICU at 24 hours and 2193 had both heart rate and SpO2 data. The final development cohort included 336 subsequent deaths (15.3%). eICU contained 364 first-unit stays; 289 patients remained in ICU and hospital at 24 hours and 287 had complete heart rate and SpO2. The external cohort contained 35 deaths (12.2%) from 65 hospitals (Fig. 1).

Median age was 71 years in MIMIC-IV and 69 years in eICU, and 64.5% and 59.6% were men, respectively. Median creatinine was 1.2 and 1.3 mg/dL; mean first-day heart rate had a median of 82.8 beats/min in both cohorts (Table 1). Only four MIMIC-IV patients (0.2%) lacked creatinine. By contrast, liver-test missingness was 46.6%-67.1% in MIMIC-IV and 42.5%-45.6% in eICU; SBP change was missing in 45.8% and 10.8%, respectively (Supplementary Table S1).

Table 1
Landmark cohort characteristics

Characteristic	MIMIC-IV	eICU
Participants, n	2193	287
Subsequent deaths, n	336	35
Age, years	71.0 (63.0–78.0)	69.0 (61.0–76.0)
Male sex, n (%)	1414 (64.5%)	171 (59.6%)
Creatinine, mg/dL	1.2 (0.9-2.0)	1.3 (0.9–2.2)
Mean SpO2, %	97.2 (95.9–98.4)	96.8 (95.3–98.3)
Mean heart rate, beats/min	82.8 (73.3–91.7)	82.8 (70.7–93.0)
Delta SBP, mmHg	-2.5 (-12.8-7.9)	-1.1 (-12.9-10.2)
ALT, U/L	31.0 (18.5–67.0)	29.5 (19.0-53.8)
AST, U/L	54.0 (29.0-142.5)	44.0 (26.0-103.5)
Albumin, g/dL	3.3 (2.9–3.7)	3.2 (2.8–3.7)
Total bilirubin, mg/dL	0.6 (0.4–0.9)	0.6 (0.4–0.9)
Alkaline phosphatase, U/L	84.0 (63.0-116.0)	79.5 (59.0-105.8)

Values are median (interquartile range) unless otherwise stated. Delta SBP is mean SBP at 18–24 h minus mean SBP at 0–6 h and is shown only descriptively; it was not a primary-model predictor.

Development and external validation

In MIMIC-IV, the apparent AUROC was 0.666 for the reference model and 0.695 for the heart-rate-augmented model. After 500 bootstrap resamples, optimism-corrected AUROC values were 0.660 and 0.690, respectively. Corresponding optimism-corrected AUPRC values were 0.243 and 0.291, and Brier scores were 0.125 and 0.122 (Table 2).

Table 2
Internal and external model performance

Dataset	Model	n/events	AUROC	AUPRC	Brier	Cal. intercept	Cal. slope
MIMIC-IV	Reference	2193/336	0.660	0.243	0.125	0.001	0.967
MIMIC-IV	Heart-rate-augmented	2193/336	0.690	0.291	0.122	0.001	0.975
eICU	Reference	287/35	0.691 (0.623–0.798)	0.216 (0.146–0.344)	0.103 (0.079–0.122)	-0.299 (-0.699–0.027)	1.100 (0.655–1.978)
eICU	Heart-rate-augmented	287/35	0.742 (0.671–0.849)	0.284 (0.179–0.491)	0.099 (0.074–0.117)	-0.303 (-0.676–0.044)	1.187 (0.746–2.112)

MIMIC-IV values are optimism-corrected estimates from 500 bootstrap resamples. eICU parentheses are 95% confidence intervals from 1000 hospital-cluster bootstrap resamples.

In eICU, the reference model achieved an AUROC of 0.691 (hospital-cluster bootstrap 95% CI 0.623–0.798), whereas the augmented model achieved 0.742 (95% CI 0.671–0.849). The paired AUROC difference was 0.051 (95% CI 0.001-0.100). AUPRC increased from 0.216 to 0.284, but its paired difference of 0.068 had a 95% CI from − 0.027 to 0.206. Brier score decreased from 0.103 to 0.099; the paired change was − 0.0045 (95% CI -0.0116 to 0.0017). Thus, only the AUROC difference excluded zero (Table 2; Fig. 2a-b).

The augmented model had an external calibration intercept of -0.303 (95% CI -0.676 to -0.044) and slope of 1.187 (95% CI 0.746–2.112). Grouped calibration showed general risk ordering but wide intervals in the highest-risk groups. Decision curves suggested a small net-benefit advantage over the reference model across parts of the 0.05–0.30 threshold range, with variation near the upper thresholds (Fig. 2c-d).

Predictor effects and sensitivity analyses

In the augmented MIMIC-IV model, the odds ratio (OR) for mean heart rate was 1.329 per 10 beats/min (95% CI 1.222–1.445). Age (OR 1.426 per 10 years, 95% CI 1.271-1.600) and creatinine (OR 1.644 per doubling, 95% CI 1.460–1.852) were also associated with subsequent mortality. Female sex and mean SpO2 were not independently associated after adjustment (Table 3; Fig. 3).

Table 3
Heart-rate-augmented model coefficients

Term	Beta	Odds ratio (95% CI)	P value
Intercept	-6.593879	-	0.046
Age, per 10 years	0.354741	1.426 (1.271-1.600)	< 0.001
Female sex	0.134148	1.144 (0.888–1.473)	0.299
Creatinine, per doubling	0.497314	1.644 (1.460–1.852)	< 0.001
Mean SpO2, per 1%	-0.004665	0.995 (0.933–1.062)	0.887
Mean heart rate, per 10 beats/min	0.284407	1.329 (1.222–1.445)	< 0.001

Coefficients were estimated in MIMIC-IV. The complete linear predictor and frozen preprocessing parameters are reported in Methods and Supplementary Table S2.

The liver-expanded augmented model yielded an external AUROC of 0.733 and AUPRC of 0.323, but depended on extensive median imputation and was therefore not promoted to the primary model. In the SBP-complete cohort (MIMIC-IV n = 1189; eICU n = 256), adding SBP change alone reduced external AUROC from 0.690 to 0.681; its paired 95% CI for AUROC change was − 0.034 to 0.018. Heart rate remained the informative dynamic feature (Table 4).

Table 4
External sensitivity analyses and APACHE IVa benchmark

Analysis	Model	n/events	AUROC	AUPRC	Brier
Liver-expanded	Expanded reference	287/35	0.691	0.265	0.101
Liver-expanded	Expanded + heart rate	287/35	0.733	0.323	0.097
SBP-complete	Reference	256/28	0.690	0.219	0.094
SBP-complete	Heart-rate-augmented	256/28	0.773	0.278	0.089
SBP-complete	Delta-SBP-augmented	256/28	0.681	0.224	0.094
SBP-complete	Heart-rate + delta-SBP	256/28	0.768	0.283	0.089
APACHE common subset	APACHE IVa	269/30	0.809	0.484	0.108
APACHE common subset	Heart-rate-augmented	269/30	0.739	0.271	0.093

Sensitivity analyses are not substitutes for the prespecified primary comparison. Liver-expanded models used development-cohort median imputation. APACHE IVa was available only in eICU.

On the APACHE IVa common subset of 269 eICU patients with 30 deaths, the augmented model AUROC was 0.739 and APACHE IVa AUROC was 0.809. The cluster-bootstrap 95% CI for their difference (augmented minus APACHE IVa) was − 0.172 to 0.048. APACHE IVa had AUPRC 0.484 versus 0.271, while the augmented model had a lower Brier score (0.093 versus 0.108); all paired intervals were imprecise. Category-based NRI was 0.271 (95% CI 0.086–0.504) and IDI 0.040 (95% CI 0.007–0.075), but these sensitivity metrics did not alter the primary interpretation.

Discussion

In this two-database 24-hour landmark study of critically ill adults with type 2 diabetes and AMI, first-day mean heart rate produced a modest but externally supported increase in discrimination beyond age, sex, creatinine, and mean SpO2. The most defensible evidence was the paired AUROC increase of 0.051, whose hospital-cluster bootstrap interval narrowly excluded zero. Improvements in AUPRC and Brier score were directionally favourable but their intervals included no difference. This distinction is important because only 35 deaths occurred in the external cohort.

The finding is biologically plausible. Diabetes may combine autonomic dysfunction, microvascular disease, renal impairment, and a higher burden of heart failure or shock [2, 19]. Mean heart rate over 24 hours can integrate sustained sympathetic activation, arrhythmia, treatment response, pain, infection, and haemodynamic compromise. It should not be interpreted as a causal exposure or a treatment target from this analysis. The null contribution of SBP change is equally informative: a two-window difference may be noisy, treatment-dependent, and available only in a selected subset.

The landmark design changes the clinical meaning of the model. It is not an admission triage score and cannot inform decisions during the first 24 hours. Instead, it estimates residual in-hospital risk among patients who have survived and remained under intensive care through the first day. This temporal separation addresses a major weakness of the previous analysis and aligns the predictor window with the intended reassessment moment.

Our APACHE IVa comparison provides an important boundary. The compact augmented model did not demonstrate superior discrimination on the common eICU subset, and APACHE IVa had higher point estimates for AUROC and AUPRC [19, 20]. The compact model uses fewer routinely available inputs and had a lower Brier score in this sample, but these observations do not establish clinical advantage. A fair comparison with GRACE was not feasible because several required elements were not consistently available across the two ICU databases [4].

The analysis also illustrates why feature availability should precede biological enthusiasm. Liver measures were clinically plausible and the expanded model performed similarly, but 42%-67% missingness made extensive imputation unavoidable. By keeping them outside the primary model, we reduced dependence on local ordering practices. Recent multi-database studies have similarly gained strength from explicit temporal definitions, independent validation, and interpretable sensitivity analyses [8].

Potential use, if confirmed prospectively, would be limited to first-day risk reassessment and research stratification. Calibration-in-the-large was negative in eICU, indicating that MIMIC-IV-derived risks tended to be too high on average. Local recalibration, clinical-impact evaluation, and predefined action pathways would therefore be required before any bedside implementation.

Strengths and limitations

Strengths include a temporally valid landmark, independent multicentre validation, frozen cross-database preprocessing, hospital-cluster uncertainty, internal bootstrap correction, calibration and decision-curve assessment, an APACHE IVa benchmark, complete coefficient reporting, and explicit negative sensitivity results.

Limitations are substantial. The retrospective phenotype depends on codes and diagnosis strings rather than adjudicated AMI and diabetes; treatments during the first 24 hours may influence both predictors and outcome; only patients remaining in ICU at 24 hours are represented; eICU contained 35 deaths and the APACHE subset 30 deaths; both databases reflect US critical care from different eras; missingness may be informative; mean values discard within-day trajectories; hospital-level validation cannot substitute for prospective evaluation; and GRACE could not be reconstructed consistently. The earlier chest-pain question cannot be answered with these fields and requires a source with structured symptom documentation or chart adjudication.

Conclusions

Among adults with type 2 diabetes and AMI who remained in intensive care at 24 hours, first-day mean heart rate modestly improved externally validated prediction of subsequent in-hospital mortality. SBP change added no stable value, and the compact model did not outperform APACHE IVa. Prospective, locally recalibrated evaluation is required before clinical use.

Abbreviations

AMI
acute myocardial infarction
APACHE
Acute Physiology and Chronic Health Evaluation
AUPRC
area under the precision-recall curve
AUROC
area under the receiver operating characteristic curve
CI
confidence interval
EHR

electronic health record
eICU
eICU Collaborative Research Database
IDI
integrated discrimination improvement
ICU
intensive care unit
MIMIC-IV
Medical Information Mart for Intensive Care IV
NRI
net reclassification improvement
OR
odds ratio
SBP
systolic blood pressure
SpO2
peripheral oxygen saturation
TRIPOD
Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.

Declarations

Ethics approval and consent to participate

MIMIC-IV and eICU contain deidentified data released under their respective institutional approvals and data-use agreements. Individual consent for this secondary analysis was not required by the source databases.

Consent for publication

Not applicable.

Competing interests

There is no conflict of interest

Use of generative artificial intelligence

This study utilized the OpenAI large language model to assist in code review. All cohort definitions, analyses, numerical results, citations, and interpretations were reviewed and verified by the authors, who take full responsibility for the manuscript.

Funding

1.Supported by Health Commission of Sichuan Province Medical Science and Technology Program(No: 24LCYJPT19)

2.Sichuan Provincial Department of Science and Technology Natural Science Foundation General Program (No: 2025ZNSFSC0707)

3.Luzhou Science and Technology Plan Project (No: 2025JYJ024; 2025JYJ087)

4.Key Laboratory of Medical Electrophysiology Open Fund research project(No.: KeyME- 2023-01)

Author Contribution

Y.FX and N.YM conceived the study. T.HW designed the phenotype and extracted the data. L.Z performed the statistical analysis. C.Z and H. HZ interpreted the clinical findings. Y.T and L.L drafted the manuscript. All authors revised the manuscript critically and approved the submitted version.

Acknowledgements

The authors thank the investigators and participants whose data contributed to MIMIC-IV and eICU and the PhysioNet team for maintaining credentialed access.

Availability of data and materials

MIMIC-IV and eICU are available to credentialed researchers through PhysioNet under their data-use agreements. Patient-level records cannot be redistributed. Analysis code, cohort definitions, preprocessing parameters, aggregate table data, and figure-source data will be deposited.

References

1. Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA, et al. Fourth universal definition of myocardial infarction (2018). Circulation. 2018;138:e618–51. 10.1161/CIR.0000000000000617.
2. Cosentino F, Grant PJ, Aboyans V, Bailey CJ, Ceriello A, Delgado V, et al. 2019 ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. Eur Heart J. 2020;41:255–323. 10.1093/eurheartj/ehz486.
3. Canto JG, Shlipak MG, Rogers WJ, Malmgren JA, Frederick PD, Lambrew CT, et al. Prevalence, clinical characteristics, and mortality among patients with myocardial infarction presenting without chest pain. JAMA. 2000;283:3223–9. 10.1001/jama.283.24.3223.

4. Fox KAA, Dabbous OH, Goldberg RJ, Pieper KS, Eagle KA, Van de Werf F, et al. Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome. BMJ. 2006;333:1091. 10.1136/bmj.38985.646481.55.
5. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med. 2006;34:1297–310. 10.1097/01.CCM.0000215112.84523.F0.
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. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H, et al. The SOFA score to describe organ dysfunction/failure. Intensive Care Med. 1996;22:707–10. 10.1007/BF01709751.
8. Ding XY, Xu HP, Zhang JR, Chen H. Temporal dimensions of oxygen saturation exposure and clinical outcomes in critically ill patients: a retrospective multi-database cohort study. Crit Care. 2025;29:402. 10.1186/s13054-025-05652-9.
9. 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. 10.1038/s41597-022-01899-x.
10. 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.
11. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet. Circulation. 2000;101:e215–20. 10.1161/01.CIR.101.23.e215.
12. 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.
13. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170:51–8. 10.7326/M18-1376.
14. Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;368:m441. 10.1136/bmj.m441.
15. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17:230. 10.1186/s12916-019-1466-7.
16. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26:565–74. 10.1177/0272989X06295361.
17. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016;69:245–7. 10.1016/j.jclinepi.2015.04.005.
18. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves. Biometrics. 1988;44:837–45. 10.2307/2531595.

19. Cheng J, Cheng Q, Wu Y, Yin J, He F. Association between triglyceride-glucose-body mass index and adverse prognosis in elderly patients with severe heart failure and type 2 diabetes. Cardiovasc Diabetol. 2025;24:299. 10.1186/s12933-025-02870-x.
20. Pencina MJ, D'Agostino RB, Sr RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker. Stat Med. 2008;27:157–172. 10.1002/sim.2929

Figures

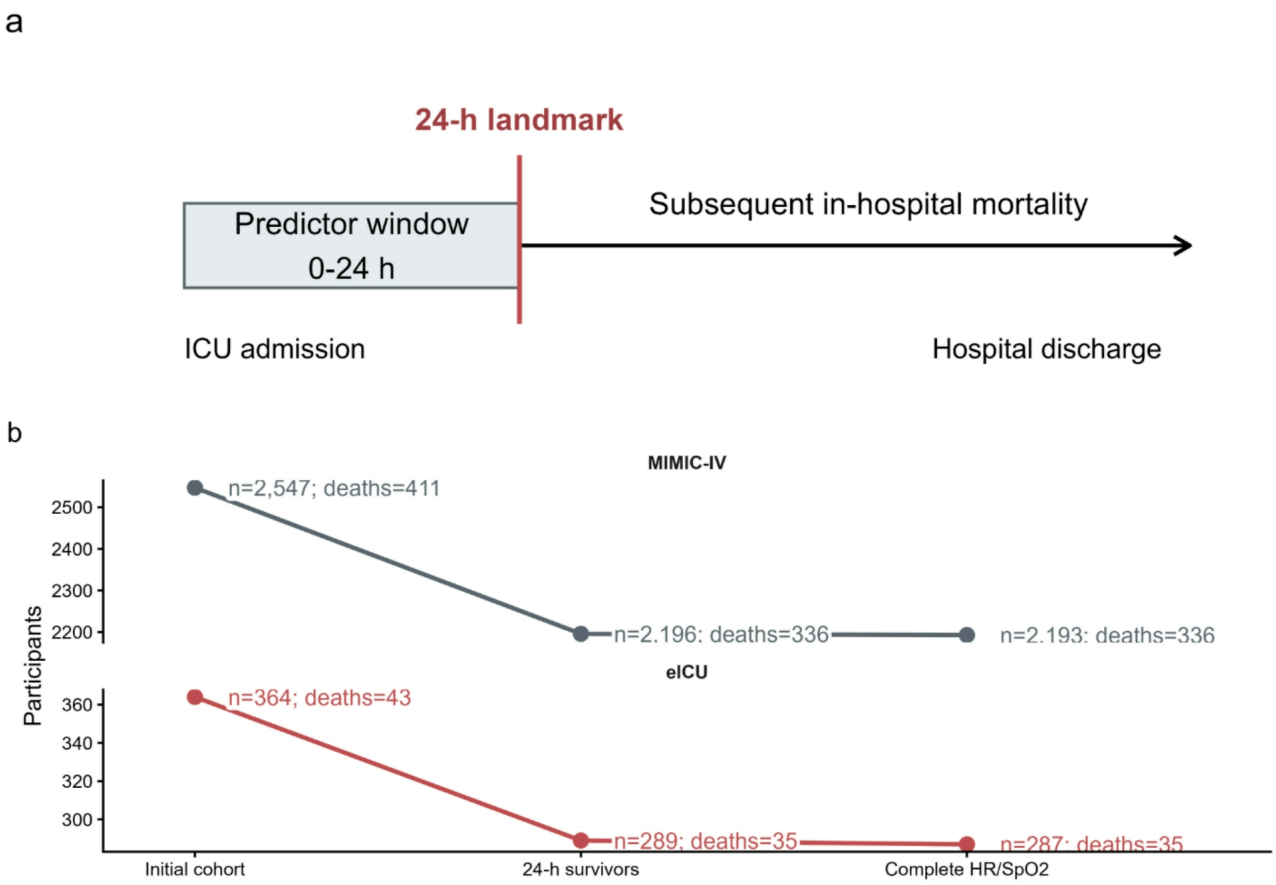


Figure 1

Landmark design and cohort derivation. Panel a shows the 0-24-hour predictor window, the 24-hour landmark, and the subsequent outcome period. Panel b shows attrition in MIMIC-IV and eICU. HR, heart rate; SpO2, peripheral oxygen saturation.

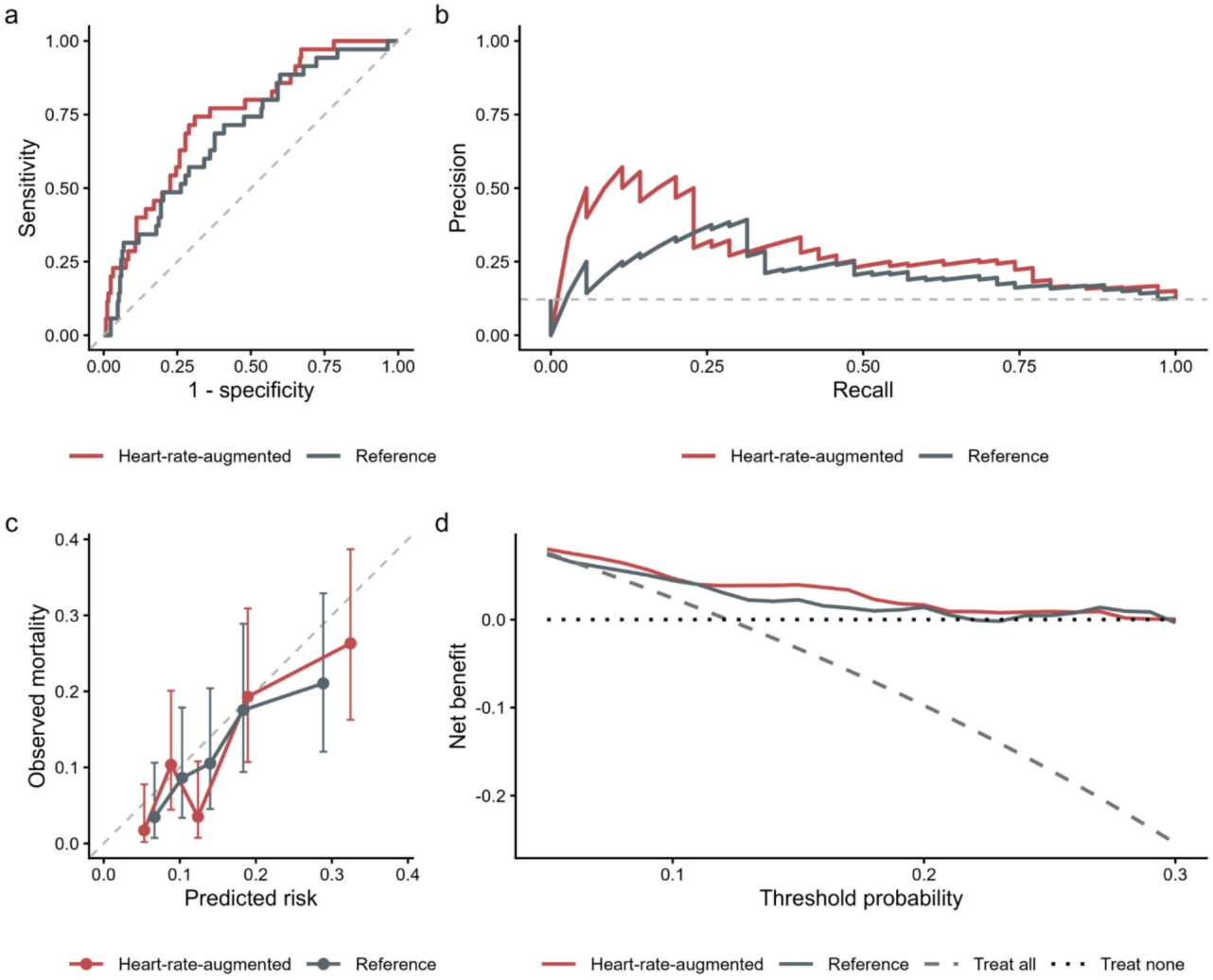


Figure 2

External validation in eICU. Receiver operating characteristic curves (a), precision-recall curves (b), grouped calibration with 95% beta-binomial intervals (c), and decision-curve analysis (d). Models were estimated in MIMIC-IV and applied without refitting. The dashed precision-recall reference is outcome prevalence.

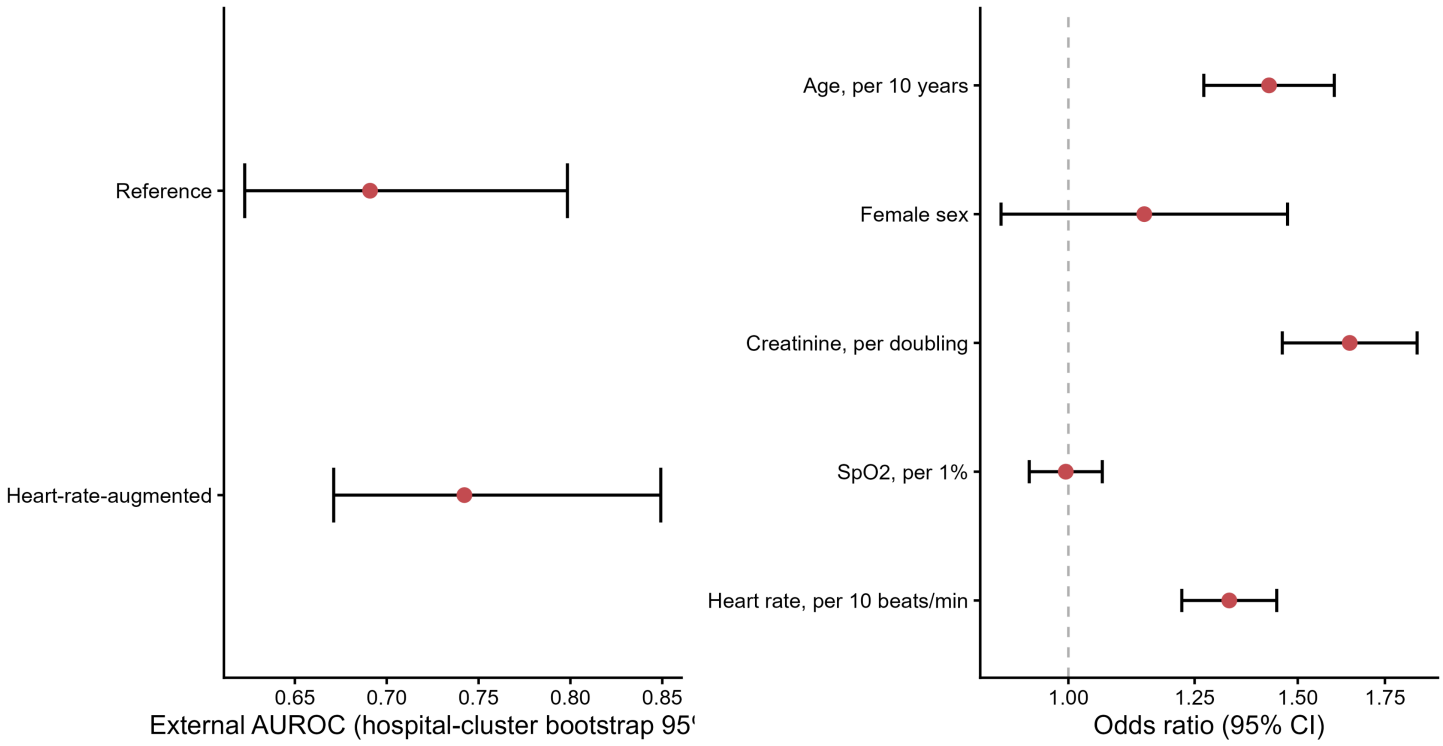


Figure 3

External discrimination and model coefficients. Panel a shows hospital-cluster bootstrap 95% confidence intervals for external AUROC. Panel b shows MIMIC-IV odds ratios with Wald 95% confidence intervals for the heart-rate-augmented model.

Supplementary Files

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

- Graphicalabstract.docx
- SupplementaryTables.docx