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.

Early potassium results for predicting hypokalemia after furosemide in critically ill adults with cardiovascular diagnoses: a multicohort study

Shuting Kong[1]
Jinhua Central Hospital
Yibin Pan[1]
zqwen628@163.com

Jinhua Central Hospital

Research Article

Keywords: furosemide, hypokalemia, intensive care, potassium, prediction model, electronic medication administration, MIMIC-IV

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11631569/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

Hypokalemia after furosemide can increase the risk of ventricular arrhythmias in susceptible patients. Earlier prediction studies have not established whether a potassium result available by 12 hours improves prediction of later low potassium or whether that gain holds across hospitals.

Methods

In a retrospective MIMIC-IV cohort, we compared a baseline logistic model with one adding potassium available by a 12-hour landmark after furosemide. Later hypokalemia was potassium < 3.5 mmol/L from 12 to < 48 hours; non-events required a late test. We assessed the models in later same-hospital records and applied separately fitted, fixed dose-free models to NWICU and INSPIRE using proxy laboratory times. Paired DeLong tests compared assessment-cohort areas under the receiver operating characteristic curve (AUCs); patient-level bootstrap resampling provided 95% intervals.

Results

Of 2,840 MIMIC-IV patients with cardiovascular discharge diagnoses, 413 had later recorded hypokalemia. Adding early potassium raised the optimism-corrected AUC from 0.686 to 0.745 and lowered the Brier score from 0.1172 to 0.1121. The AUC gain in 953 later same-hospital patients was 0.054 (95% bootstrap interval 0.019–0.089). It was 0.092 in NWICU (n = 369; 65 events; 95% interval 0.018–0.169; P = 0.018), but uncertain in INSPIRE cardiovascular patients (n = 132; 31 events; 0.020, 95% interval − 0.096 to 0.134; P = 0.737). A broader INSPIRE postoperative ICU cohort had a gain of 0.097 (n = 247; 69 events; 95% interval 0.021–0.168; P = 0.012), with underpredicted absolute risk. In MIMIC-IV, 19.2% with an early result lacked enough later testing to classify the outcome.

Conclusions

Early potassium improved discrimination in selected MIMIC-IV patients. The gain was uncertain in the INSPIRE cardiovascular subgroup, and the original model lacks independent external validation. Prospective testing is needed before using it to guide monitoring.

Trial registration

Not applicable. This was an observational prediction-model study and no clinical trial was conducted.

Background

Furosemide is often needed to treat congestion [1, 2], but low potassium can destabilize cardiac repolarization and contribute to ventricular tachycardia or fibrillation in susceptible patients [3, 4]. The practical question is whether an early potassium result helps identify patients likely to have a later low result. This study measured laboratory hypokalemia, not arrhythmias or sudden death.

At 12 hours, a pre-treatment sample may be old or may have returned only after furosemide was given. Diuresis, renal handling, potassium replacement, and other care can also change potassium during those hours. We therefore compared models using pre-index and new results at the same 12-hour decision point. “Baseline” describes when a sample was drawn, not when clinicians first saw its result.

An existing furosemide score studied acute decompensated heart failure [5], and a broader inpatient model predicted a more severe potassium endpoint [6]. Neither quantified the added value of a new result at a fixed 12-hour decision point while requiring enough later testing to identify non-events. A related 12-hour loop-diuretic study examined kidney deterioration [7].

We compared baseline and updated models in the same MIMIC-IV patients, assessed them in later records from that hospital, and tested separate fixed dose-free models in NWICU and INSPIRE. The study asks whether an early result improves prediction of a later laboratory result; it does not estimate a causal furosemide effect or the benefit of acting on a score.

Methods

Data source and design

We developed prediction models in a retrospective MIMIC-IV v3.1 cohort [8, 9], distributed through PhysioNet [10], and assessed them in two independent hospital datasets. The hospital module supplied admissions, diagnoses, prescriptions, electronic medication administration records (eMAR), and laboratory results; the intensive care unit (ICU) module supplied stay times. We used the original source tables to rebuild eligibility, treatment-record timing, and laboratory availability. Patient-level data remained in the credentialed local environment. Reporting was guided by TRIPOD + AI [11].

For laboratory records missing a hospital admission identifier, we retained a result only when the patient identifier matched and its sample time fell unambiguously within the index admission; the MIMIC-IV documentation permits missing admission identifiers in this table [12]. We made this source-field correction after viewing initial overall results and retained aggregate outputs from both versions. We also corrected the handling of age top-coding after viewing initial overall results, as described below. The final results here use the corrected definitions.

Cohort and index record

Adults aged at least 18 years were eligible if their index hospital admission carried an International Classification of Diseases, Tenth Revision (ICD-10) diagnosis in I20–I52 or a Ninth Revision (ICD-9) diagnosis in 410–429. These discharge codes define a broad, retrospective cardiovascular-diagnosis cohort. They do not show that a cardiac diagnosis was known at 12 hours or caused ICU admission. If a patient had several eligible ICU stays, we selected the earliest qualifying event for that patient.

The index was the first furosemide (Lasix) eMAR entry within 24 hours of ICU admission marked Administered. We required a linked intravenous (IV) prescription matched by admission and pharmacy identifier, plus a positive dose or product amount in an eMAR detail row without a not-given flag. The eMAR charttime set time zero (t0). Its documentation storetime had to be no later than t0 + 12 hours. This was the first *qualifying ICU record*, which could follow another furosemide administration during the same hospital admission. The linked prescription identifies the ordered IV route; an independent bedside route field was not available in the selected detail records. The recorded index dose is the amount for that event, not cumulative delivered dose.

Landmark, predictors, and outcome

Patients had to be alive, hospitalized, and in the ICU at t0 + 12 hours. Potassium was a numeric blood chemistry result for MIMIC item 50971 in the range 1.5–8 mmol/L. Baseline potassium was the latest sample from t0 − 24 hours to before t0 whose result was available by the landmark. It had to be at least 3.5 mmol/L. We excluded patients with any valid low potassium available from t0 − 24 hours to before the landmark and those with a potassium sample from that period still pending at the landmark. For the dynamic comparison, a new potassium sample from t0 to before t0 + 12 hours also had to have its result returned by the landmark. We did not assign a change of zero to patients without a new sample. Result time and event identifier resolved tied sample times.

The baseline clinical model used baseline potassium, age, recorded sex, the last pre-index creatinine (item 50912) available by the landmark, and the recorded index furosemide dose. Age was calculated from MIMIC anchor age and shifted admission year and capped at 91 years because anchor age is top-coded for patients older than 89 [13]. Creatinine and dose were log2-transformed. Creatinine is shown in µmol/L; model fitting used the source numerical scale, and Additional file 1 gives the equivalent SI-unit input rule. The updated model added the latest new potassium result available by 12 hours. A model containing baseline potassium alone served as a simple reference. We chose predictors for clinical relevance, parsimony, and availability at the landmark, without selection by univariable P values. Later medication, later laboratory frequency, potassium replacement, urine output, and discharge diagnoses were not predictors.

The outcome was any valid potassium < 3.5 mmol/L sampled from t0 + 12 to < 48 hours, before hospital discharge or death in the same admission. Without a low result, a patient was classified as a non-event only if a valid potassium was sampled from t0 + 36 to < 48 hours before discharge or death. Otherwise the outcome was unknown. Outcome timing used sample charttime; predictor availability used result

storetime. A measurement after ICU transfer but during the index admission could establish the outcome. Both models were compared for use at t0 + 12 hours. “Baseline” means sampled before the index; its result was not required to be available before the index.

Statistical analysis

This retrospective cohort study used a fixed 12-hour landmark and a binary, laboratory-recorded outcome. Age was summarized as mean (standard deviation), other continuous variables as median (interquartile range), and categorical variables as number (percentage). For the descriptive outcome-group comparisons in Table 1, we used a two-sided Welch t test for age, Mann–Whitney U tests for other continuous variables, and Pearson χ^2 tests for categorical variables; no expected cell count required Fisher's exact test. P values are unadjusted descriptions, not criteria for predictor selection.

Three logistic models were fitted to the same 2,840 patients with an early potassium result and classifiable outcome. Predictors were chosen for clinical relevance and landmark availability, without univariable P-value screening. Two patients lacked baseline creatinine; median imputation and continuous-variable standardization were fitted within training data. Performance was assessed by the area under the receiver operating characteristic curve (AUC), average precision, Brier score, and calibration [14, 15]. Calibration-in-the-large was the log-odds intercept with the calibration slope fixed at one. Three hundred patient-level bootstrap samples refitted the complete pipeline for optimism correction. The reported bootstrap interval for the primary apparent AUC difference does not describe uncertainty in the optimism-corrected estimate. The fitted six-variable updated model was rendered as a research-only nomogram [16]. The complete raw-scale equations, predictor coding, transformations, and imputation values for all five fitted models are given in Additional file 1.

For an internal temporal assessment, MIMIC anchor-year fields defined nonoverlapping earlier and later possible calendar-year intervals [8]. Earlier records supplied all fitting steps; later records were assessed without refitting or recalibration. Two thousand patient-level resamples of the later set provided 95% percentile intervals. We used paired, two-sided DeLong tests to compare AUCs for fixed predictions in the same patients in the later MIMIC, NWICU, and INSPIRE cohorts [17].

We screened local eICU [18] and NWICU [19] fields before cross-database analysis. The eICU cardiac-unit subset had only 22 classifiable patients and 7 events, with continuous-infusion records and ICU-only laboratory follow-up. For exploratory NWICU external validation, we fixed a reduced comparison before viewing its performance because no comparable administration dose was available. Four-variable baseline and five-variable 12-hour-updated models were newly trained in all 2,840 MIMIC patients. A first NWICU ICU furosemide eMAR entry marked Applied with an INJECTION prescription within 24 hours served as a proxy index. We applied the 12-hour risk and 12-to-<48-hour outcome rules where fields permitted. MIMIC imputation, scaling, and coefficients were applied unchanged to NWICU; 2,000 patient bootstrap resamples supplied 95% percentile intervals. NWICU Applied, INJECTION, and laboratory storetime are not proven equivalent to MIMIC Administered, IV route, and result time. NWICU storetime

was treated as a recorded-time proxy for availability, not proof that clinicians could see a result at 12 hours. This is an **exploratory external validation of newly fitted dose-free models under a proxy definition**, not validation of the original six-variable model or of real-time use.

We also assessed INSPIRE v1.4.2 [20, 21], an independent South Korean perioperative database. Its medication table records administration and IV route, but has no comparable dose; its laboratory table gives sample times without a separate result-availability time. The primary INSPIRE cohort required an adult postoperative ICU stay, an I20–I52 diagnosis recorded within that admission, and the first IV furosemide administration during the first 24 ICU hours; the earliest qualifying admission was retained per patient. At 12 hours patients had to remain in ICU, alive, and hospitalized. We applied the same 24-hour pre-index baseline, 0-to-<12-hour new potassium, and 12-to-<48-hour outcome windows, but used sample time as an **availability proxy**. Patients without a low result required a potassium test at 36 to < 48 hours to classify a non-event. Laboratory values were restricted to the index admission; same-minute duplicate values were averaged. Age was capped at 91 years, and the MIMIC-trained creatinine imputation and dose-free coefficients were applied without refitting or recalibration. As a secondary transportability analysis, we repeated this assessment in the broader postoperative ICU population without requiring a cardiovascular diagnosis. Two thousand patient bootstrap resamples supplied percentile intervals. INSPIRE rounds laboratory values to 5th-percentile bins [20], which can affect the < 3.5 mmol/L threshold and discrimination; its sampling time does not prove that the result was visible at 12 hours.

Sensitivity analyses excluded earlier same-admission furosemide, excluded end-stage kidney disease codes, or required a late recheck for all classifiable patients. Additional checks restricted medication label and early-sample timing. A post hoc NWICU check restricted early potassium to records with storetime later than charttime; this does not prove when results became visible. Observation weighting was exploratory and relied on a missing-at-random assumption. All reported P values were two-sided; P < 0.05 was used as a descriptive significance threshold without multiplicity adjustment. Analyses used DuckDB and Python with NumPy, SciPy, and scikit-learn.

Use of a large language model

OpenAI Codex assisted with analysis and plotting code, figure layout, and manuscript drafting, translation, and language editing. Figures were drawn from verifiable aggregate analysis outputs or fitted coefficients, not generated as synthetic clinical data. The authors are responsible for reviewing the code, figures, results, citations, and final text before submission. The large language model was not an author.

Results

Cohort and measurement process

Among 41,625 adults with an admission carrying a cardiovascular diagnosis and 60,338 ICU stays, 8,711 patients had a furosemide (Lasix) eMAR entry within the first 24 ICU hours. A qualifying index was found in 7,791 patients. At 12 hours, 5,875 remained alive in the ICU and 4,260 met the identifiable non-low potassium risk-set definition. Later outcomes could be classified for 3,471 patients: 497 had recorded hypokalemia and 2,974 had an observed non-event. The main dynamic comparison included 2,840 patients with an early potassium result and a classifiable outcome, including 413 events (Fig. 1).

Of the 3,516 risk-set patients with an early potassium result, 676 (19.2%) lacked enough later testing to classify the outcome. Across the broader 4,260-patient risk set, 789 (18.5%) had an unknown outcome; 706 were still hospitalized at 36 hours. The 14.5% event fraction in the 2,840-patient analysis sample therefore describes a selected, monitored group rather than all treated patients. Eighty-one of 497 classifiable events (16.3%) in the broader risk set were first found after ICU transfer. An earlier furosemide eMAR administration in the same admission preceded the index for 2,146 of 7,791 patients (27.5%).

The latest early potassium sample was collected a median of 6.3 hours (interquartile range 4.6–8.3) after the index, leaving a median sample age of 5.7 hours at the 12-hour landmark. For 50 patients the latest early sample was within one hour of the index. Among 413 events in the dynamic sample, first low potassium was sampled a median of 29.4 hours after the index: 135 at 12–<24 hours, 151 at 24–<36 hours, and 127 at 36–<48 hours.

Patient characteristics

Patients with later recorded hypokalemia had lower baseline and early potassium and a higher recorded index dose (Table 1). The unadjusted group-comparison P values describe this selected sample and were not used to select model predictors or infer a drug effect. The heart-failure discharge-code row describes case mix; it was not a predictor.

Table 1
Characteristics of the dynamic, outcome-classifiable sample

Characteristic	No recorded hypokalemia (n = 2427)	Recorded hypokalemia (n = 413)	P value
Age, mean (SD), years; capped at 91	69.3 (12.4)	70.3 (14.2)	0.168
Male, n (%)	1636 (67.4)	225 (54.5)	< 0.001
Baseline potassium, median (IQR), mmol/L	4.5 (4.2–4.9)	4.3 (3.9–4.6)	< 0.001
Latest early potassium, median (IQR), mmol/L	4.4 (4.1–4.8)	4.0 (3.8–4.3)	< 0.001
Early minus baseline potassium, median (IQR), mmol/L	−0.1 (− 0.4 to 0.2)	−0.2 (− 0.5 to 0.1)	< 0.001
Baseline creatinine, median (IQR), µmol/L	97 (71–159)	106 (71–159)	0.071
Recorded index furosemide dose, median (IQR), mg	20 (20–40)	40 (20–80)	< 0.001
Heart-failure discharge code, n (%)	1320 (54.4)	302 (73.1)	< 0.001

Table note

SD, standard deviation; IQR, interquartile range. Age used a two-sided Welch t test; other continuous variables used two-sided Mann–Whitney U tests; categorical variables used two-sided Pearson χ^2 tests without Yates correction. P values were not adjusted for multiple comparisons. Creatinine was available in 2,426 non-event and 412 event patients; displayed SI values are converted from source mg/dL and rounded. The heart-failure code was not a predictor.

Model performance

Adding early potassium raised optimism-corrected AUC from 0.686 to 0.745, an increment of 0.058, and lowered the Brier score by 0.0051 (Table 2). Average precision also improved. The paired bootstrap 95% percentile interval for the *apparent* AUC increment was 0.039–0.081; it does not quantify uncertainty in the optimism-corrected increment. In the updated model's apparent risk deciles, mean predicted and observed event fractions were 39.9% and 38.4% in the highest tenth, compared with 2.4% and 4.2% in the lowest tenth.

Table 2
Same-sample logistic model performance after 300 patient-level bootstrap optimism correction

Model	Apparent AUC	Corrected AUC	Corrected Brier score	Corrected average precision	Corrected calibration slope
Baseline potassium only	0.624	0.625	0.1215	0.224	1.027
Baseline clinical	0.690	0.686	0.1172	0.272	0.982
Twelve-hour updated	0.748	0.745	0.1121	0.317	0.983

Table note

All models used the same 2,840 patients and 413 events. The baseline clinical model comprised baseline potassium, age, sex, baseline creatinine, and recorded index dose; the updated model added the latest early potassium. AUC, area under the receiver operating characteristic curve. Corrected estimates are point estimates; confidence intervals for those corrected metrics were not estimated. The reported AUC increment (0.058) was calculated before rounding the two displayed corrected AUCs.

Internal temporal assessment

The unambiguously earlier set comprised 1,524 patients and 232 events; the later set comprised 953 patients and 141 events. Another 363 patients and 40 events had year intervals crossing the boundary and were excluded only from this temporal assessment. Without refitting in later records, the baseline and updated models had AUCs of 0.689 (95% patient-bootstrap interval 0.637–0.733) and 0.742 (0.695–0.783), respectively (difference 0.054; two-sided patient-bootstrap 95% percentile interval 0.019–0.089; paired DeLong P = 0.004; Fig. 2A). Brier scores were 0.1193 and 0.1158, respectively (difference −0.0035, patient-bootstrap 95% interval − 0.0075 to 0.0006). The updated model predicted a mean risk of 16.1%, compared with an observed event fraction of 14.8%; its calibration slope was 0.896 (95% percentile interval 0.682–1.112) and calibration-in-the-large offset was − 0.113 (− 0.303 to 0.067; Fig. 2B). The temporal assessment supports an internal performance signal, with limited calibration precision. The research-only nomogram (Fig. 3) displays the apparent MIMIC fit; its absolute risks have not been externally validated.

Exploratory checks

The corrected AUC increment remained positive after excluding patients with an earlier same-admission furosemide record (0.078; 1,980 patients, 257 events). It was 0.056 after excluding end-stage kidney disease discharge codes (2,792 patients, 410 events) and 0.062 when a late recheck was required for both events and non-events (2,792 patients, 365 events). Checks of exact medication labels and early-

sample timing yielded increments of about 0.059–0.060. An early-potassium-only model had corrected AUC 0.710. Observation weighting gave an *apparent* AUC increment of about 0.058 under a missing-at-random assumption; it cannot account for unmeasured reasons for testing. These checks followed the main result and are exploratory.

Exploratory external validation in NWICU

In NWICU, 1,434 patients had a proxy index, 1,110 remained alive in ICU at 12 hours, and 856 met the identifiable non-low-potassium risk definition. Of 558 with an early result, 369 had a classifiable later outcome (65 events); 189 (33.9%) lacked sufficient later testing. The newly trained, dose-free models were fixed after MIMIC training and applied to these 369 NWICU patients. Adding early potassium increased AUC from 0.566 to 0.658 (difference 0.092; patient-bootstrap 95% percentile interval 0.018–0.169; paired DeLong P = 0.018; Additional file 2, Supplementary Figure S1). Brier score changed from 0.1456 to 0.1392 (difference − 0.0064; 95% interval − 0.0144 to 0.0008). The updated model's mean predicted risk was 17.63% versus 17.62% observed; calibration slope was 0.725 (95% interval 0.382–1.170; Table 3), so individual-risk calibration remains uncertain. In 1,422/1,434 (99.2%) proxy index records, eMAR storetime preceded charttime; 226/369 (61.2%) selected early potassium records had identical sample and store times. In a post hoc restriction to the other 143 patients (34 events), AUCs were 0.525 and 0.621 (difference 0.097; 95% bootstrap interval − 0.005 to 0.204). These timestamps do not verify when results were clinically visible. This exploratory external validation tests **newly fitted dose-free models in a proxy-defined population with classifiable outcomes**. It does not validate the original six-variable model, real-time use, or portability of absolute risks.

Table 3
Exploratory NWICU external validation of MIMIC-trained reduced models

Metric	Baseline, no dose	Twelve-hour updated, no dose	P value
AUC (95% bootstrap interval)	0.566 (0.490–0.640)	0.658 (0.586–0.730)	0.018
Brier score (95% bootstrap interval)	0.1456 (0.1187–0.1728)	0.1392 (0.1161–0.1628)	—
Average precision (95% bootstrap interval)	0.227 (0.159–0.318)	0.287 (0.210–0.394)	—
Mean predicted event risk	15.63%	17.63%	—
Calibration slope (95% bootstrap interval)	0.475 (− 0.025 to 1.002)	0.725 (0.382–1.170)	—
Calibration intercept, slope fixed at 1; log-odds (95% bootstrap interval)	0.148 (− 0.150 to 0.413)	−0.001 (− 0.303 to 0.268)	—

Table note: NWICU n = 369, 65 events (17.62%) among patients with sufficient early and later testing to classify the outcome; 189/558 (33.9%) with an early result had unknown later status. The AUC P value is

from a two-sided paired DeLong test of correlated ROC curves. Paired AUC difference: +0.092 (95% patient-bootstrap interval 0.018–0.169); Brier difference: −0.0064 (− 0.0144 to 0.0008). Other rows were not subjected to hypothesis tests. This proxy-defined validation applies to new reduced models without dose, not to the original six-variable model.

Exploratory external assessment in INSPIRE

In INSPIRE, 695 patients had a qualifying IV furosemide record in a postoperative ICU stay with a cardiovascular diagnosis recorded during the same admission. Of 600 still eligible at 12 hours, 179 had a new potassium sample without a recorded earlier low result. Later outcome status was classifiable in 132 (31 events); 47/179 (26.3%) had insufficient later testing. The fixed MIMIC dose-free models had AUCs of 0.504 and 0.525 after updating (difference 0.020, 95% patient-bootstrap interval − 0.096 to 0.134; paired DeLong P = 0.737; Table 4 and Additional file 3, Supplementary Figure S2A). Brier score changed from 0.1921 to 0.1946 (difference 0.0025, 95% interval − 0.0088 to 0.0133). The updated model predicted a mean risk of 13.3%, versus 23.5% observed; its calibration slope was 0.219 (95% interval − 0.278 to 0.696). This small perioperative cardiovascular subgroup did not reproduce a clear discrimination gain and absolute risk was underestimated.

In the prespecified broader postoperative ICU assessment, which includes the cardiovascular subgroup, 247/339 patients with a new potassium sample had a classifiable later outcome (69 events); 92/339 (27.1%) had unknown status. AUC increased from 0.519 to 0.616 (difference 0.097, 95% interval 0.021–0.168; P = 0.012; Table 4 and Supplementary Figure S2B). Brier score decreased from 0.2216 to 0.2112 (difference − 0.0104, 95% interval − 0.0202 to − 0.00002). Yet updated mean predicted risk was 14.2%, compared with 27.9% observed, and calibration slope was 0.593 (95% interval 0.265–0.927). The broader-group improvement cannot replace the cardiovascular subgroup result or establish portable absolute-risk estimates.

Table 4
Exploratory INSPIRE assessment of fixed MIMIC-trained dose-free models

Cohort	Baseline AUC (95% bootstrap interval)	Updated AUC (95% bootstrap interval)	Paired AUC difference (95% bootstrap interval); P value
Cardiovascular diagnosis; n = 132, 31 events	0.504 (0.388–0.617)	0.525 (0.411–0.631)	0.020 (− 0.096 to 0.134); 0.737
Broader eligible postoperative ICU; n = 247, 69 events	0.519 (0.449–0.592)	0.616 (0.539–0.689)	0.097 (0.021–0.168); 0.012

Table note

INSPIRE is an independent perioperative hospital dataset. The larger cohort includes the cardiovascular-diagnosis subgroup; the two rows are not independent validation samples. Models were fitted only in MIMIC and applied without refitting or recalibration. AUC intervals and paired differences use 2,000

patient-bootstrap resamples; P values use two-sided paired DeLong tests and are exploratory without multiplicity adjustment. Laboratory chart_time is a sample-time proxy, not a verified result-availability time; INSPIRE laboratory values are binned for deidentification. The cardiovascular subgroup had 47/179 (26.3%) and the broader group 92/339 (27.1%) with an early sample but unclassifiable later outcome. Neither row validates the original dose-containing six-variable model.

Discussion

At the 12-hour decision point, adding a new potassium result improved discrimination for later laboratory hypokalemia in monitored MIMIC patients and in later records from the same hospital. A fixed dose-free update also improved discrimination in NWICU, whereas the INSPIRE cardiovascular subgroup showed no clear gain. The broader INSPIRE cohort improved ranking but underestimated absolute risk. The value of an early result therefore needs assessment in each clinical setting.

The earlier furosemide score focused on acute decompensated heart failure [5]; a broader inpatient model predicted a different, more severe endpoint [6]. Our study compared models in the same patients at a fixed decision time and required a late potassium test before classifying a patient without a low result as a non-event. MIMIC offered medication administration records, result-availability times, and later same-hospital data for assessment without refitting. We also tested fixed dose-free models in two independent hospitals. Their different results make the limits of transportability visible. The six-variable nomogram remains a display of the MIMIC fit, not an externally validated clinical tool.

There is a plausible reason for a new result to add information. Furosemide blocks Na-K-2Cl transport in the thick ascending limb, increasing sodium delivery to the distal nephron. Distal sodium uptake favors potassium secretion; volume contraction, aldosterone activation, chloride depletion, and alkalosis can intensify potassium loss [22–25]. A pre-index potassium sample describes the starting concentration, although its result might arrive after treatment begins. The next result can reflect the combined effects of diuresis, kidney function, replacement, and other care. We did not measure those processes together and cannot isolate a causal drug effect.

Magnesium deficiency can increase renal potassium loss and make hypokalemia harder to correct [26]. Venous congestion, diuretic resistance, kidney function, and combination treatment also change the response to diuretics [27–32]. One creatinine value and one recorded dose cannot capture those changes. Perioperative fluid and acid–base shifts, insulin, catecholamines, and potassium replacement may further affect INSPIRE patients; these explanations remain untested. INSPIRE also bins laboratory values and lacks result-availability times [20, 21]. These differences may help explain why a ranking gain in the broader postoperative cohort did not yield accurate absolute risk.

The measured outcome matters because low extracellular potassium can reduce repolarizing current and Na^+/K^+ pump activity, favoring calcium accumulation and afterdepolarizations in susceptible myocardium [4, 33]. Clinical studies have linked potassium levels to ventricular arrhythmias, cardiovascular death, or mortality after acute coronary events [34–36] and in heart failure [37–40].

Reviews and the LYTE-VT study add evidence on hypokalemia and electrolyte abnormalities in arrhythmia presentations [41, 42]. Our study did not measure those events. It shows prediction of a laboratory threshold, not prevention of arrhythmia or death.

The main limitations are selective early and late testing, retrospective cardiovascular diagnoses, intervening treatment, and the absence of independent validation for the original six-variable model. A discharge diagnosis does not identify the eligible bedside population at 12 hours. NWICU medication status, route, and timestamps are proxies; 189 of 558 patients with an early result had unknown later status. INSPIRE records IV administration but only sample times, and its cardiovascular subgroup had 31 events with 47 of 179 later outcomes unknown. These constraints limit clinical use and explain why the MIMIC discrimination gain should be treated as a finding to test further.

The next study should enroll an adequately sized independent cardiovascular cohort [43] using an indication known at 12 hours, confirmed administration and cumulative dose, result-availability times, and scheduled potassium and magnesium testing through ICU transfer. It should compare perioperative and nonsurgical patients and evaluate calibration before testing a risk-based recheck schedule. A subsequent trial could then assess severe hypokalemia, ventricular arrhythmias, treatment changes, and the burden of extra blood draws.

Conclusions

In selected MIMIC patients with cardiovascular diagnoses, an early potassium result improved discrimination for later laboratory hypokalemia at a 12-hour landmark. The gain was seen in proxy-defined NWICU data but was uncertain in INSPIRE's postoperative cardiovascular subgroup. Prospective studies with comparable result times and scheduled follow-up should validate the model and test whether this information improves monitoring and patient outcomes.

Abbreviations

AUC
area under the receiver operating characteristic curve
eMAR
electronic medication administration record
ICD
International Classification of Diseases
ICU
intensive care unit
IQR
interquartile range
IV
intravenous

MIMIC-IV
Medical Information Mart for Intensive Care IV
NWICU
Northwestern ICU
ROC
receiver operating characteristic
SD
standard deviation.

Declarations

Ethics approval and consent to participate

MIMIC-IV creation and data sharing were reviewed by the Beth Israel Deaconess Medical Center Institutional Review Board, with a waiver of individual informed consent [8]. NWICU participation and deidentified data release were approved by the relevant source institutions; the Northwestern University Institutional Review Board waived individual informed consent [19]. INSPIRE data release was approved by the Seoul National University Hospital Institutional Review Board, which waived written informed consent for its anonymized source data [20]. This secondary analysis used existing deidentified data through credentialed access.

Consent for publication

Not applicable. This manuscript contains no individual participant details, images, or videos.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Jinhua Traditional Chinese Medicine Science and Technology Research Program (Project No. 2023KY05; grant recipient: Shuting Kong). The funder had no role in the study design, data analysis, manuscript preparation, or decision to publish.

Author Contribution

SK contributed to study conceptualization, data analysis and interpretation, manuscript revision, and funding acquisition. YP contributed to clinical interpretation, supervision, and critical manuscript revision. Both authors read and approved the final manuscript.

Acknowledgement

We thank the patients whose deidentified health records made this secondary analysis possible, and the MIMIC-IV, NWICU, INSPIRE, and PhysioNet teams for creating and maintaining these research resources.

Data Availability

MIMIC-IV v3.1, NWICU v0.1.0, and INSPIRE v1.4.2 are available to credentialed researchers through the PhysioNet dataset pages cited in references 8, 19, and 20, under their respective data use agreements. Patient-level source data and derived patient-level extracts cannot be redistributed. Aggregate results are reported in this article and its additional files. The analysis scripts, without patient-level data, are archived at Zenodo (version 1.0.0; https://doi.org/10.5281/zenodo.23969427). The archive requires authorized source data and private staging databases whose construction scripts are not included.

References

1. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2021;42(36):3599–726. https://doi.org/10.1093/eurheartj/ehab368.
2. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, et al. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2022;145(18):e895–1032. https://doi.org/10.1161/cir.0000000000001063.
3. Krogager ML, Kragholm K, Thomassen JQ, Søgaard P, Lewis BS, Wassmann S, et al. Update on management of hypokalaemia and goals for the lower potassium level in patients with cardiovascular disease: a review in collaboration with the European Society of Cardiology Working Group on Cardiovascular Pharmacotherapy. Eur Heart J Cardiovasc Pharmacother. 2021;7(6):557–67. https://doi.org/10.1093/ehjcvp/pvab038.
4. Skogestad J, Aronsen JM. Hypokalemia-Induced Arrhythmias and Heart Failure: New Insights and Implications for Therapy. Front Physiol. 2018;9:1500. https://doi.org/10.3389/fphys.2018.01500.
5. Rakphaka Y, Sornhiran P. Clinical prediction model for hypokalemia in hospitalized patients with acute decompensated heart failure treated with intravenous furosemide. J Nephrol Soc Thail. 2025;31(4):326–36. https://doi.org/10.63555/jnst.2025.282304.
6. Li Y, Staley B, Henriksen C, Xu D, Lipori G, Winterstein AG. Development and validation of a dynamic inpatient risk prediction model for clinically significant hypokalemia using electronic health record data. Am J Health Syst Pharm. 2019;76(5):301–11. https://doi.org/10.1093/ajhp/zxy051.
7. Wang J, Deng N, Huang Y, et al. Early loop-diuretic efficiency and subsequent worsening renal function in critically ill adults with a coded acute heart-failure diagnosis: a 12-hour landmark analysis of MIMIC-IV. BMC Cardiovasc Disord. 2026. https://doi.org/10.1186/s12872-026-06913-3.

8. Johnson A, Bulgarelli L, Pollard T, Gow B, Moody B, Horng S et al. MIMIC-IV (version 3.1). PhysioNet. 2024. https://doi.org/10.13026/kpb9-mt58. Dataset: https://physionet.org/content/mimiciv/3.1/. Accessed 5 Nov 2026.
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):1. https://doi.org/10.1038/s41597-022-01899-x.
10. Pollard T, Moody BE, Lehman L, Gow BJ, Fernandes C, Xie C, et al. PhysioNet as a global platform for biomedical research. Nat Health. 2026;1(8):792–5. https://doi.org/10.1038/s44360-026-00096-z.
11. 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. https://doi.org/10.1136/bmj-2023-078378.
12. MIMIC-IV documentation. Hospital module: labevents. https://mimic.mit.edu/docs/iv/modules/hosp/labevents.html. Accessed 5 Nov 2026.
13. MIMIC-IV documentation. Hospital module: patients. https://mimic.mit.edu/docs/iv/modules/hosp/patients.html. Accessed 5 Nov 2026.
14. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128–38. https://doi.org/10.1097/ede.0b013e3181c30fb2.
15. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, et al. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. https://doi.org/10.1186/s12916-019-1466-7.
16. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008;26(8):1364–70. https://doi.org/10.1200/jco.2007.12.9791.
17. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837–45. https://doi.org/10.2307/2531595.
18. 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. https://doi.org/10.1038/sdata.2018.178.
19. Moukheiber D, Temps W, Molgi B, Li Y, Lu A, Nannapaneni P et al. Northwestern ICU (NWICU) database (version 0.1.0). PhysioNet. 2024. https://doi.org/10.13026/s84w-1829. Dataset: https://physionet.org/content/nwicu-northwestern-icu/0.1.0/. Accessed 5 Nov 2026.
20. Lim L, Lee HC. INSPIRE, a publicly available research dataset for perioperative medicine (version 1.4.2). PhysioNet. 2026. https://doi.org/10.13026/1eay-yc85. Dataset: https://physionet.org/content/inspire/1.4.2/. Accessed 5 Nov 2026.
21. Lim L, Lee H, Jung CW, Sim D, Borrat X, Pollard TJ, et al. INSPIRE, a publicly available research dataset for perioperative medicine. Sci Data. 2024;11:655. https://doi.org/10.1038/s41597-024-03517-4.

22. Ellison DH. Clinical Pharmacology in Diuretic Use. Clin J Am Soc Nephrol. 2019;14(8):1248–57. https://doi.org/10.2215/cjn.09630818.
23. Sica DA. Diuretic-related side effects: development and treatment. J Clin Hypertens (Greenwich). 2004;6(9):532–40. https://doi.org/10.1111/j.1524-6175.2004.03789.x.
24. Palmer BF, Clegg DJ. Physiology and Pathophysiology of Potassium Homeostasis: Core Curriculum 2019. Am J Kidney Dis. 2019;74(5):682–95. https://doi.org/10.1053/j.ajkd.2019.03.427.
25. Do C, Vasquez PC, Soleimani M. Metabolic Alkalosis Pathogenesis, Diagnosis, and Treatment: Core Curriculum 2022. Am J Kidney Dis. 2022;80(4):536–51. https://doi.org/10.1053/j.ajkd.2021.12.016.
26. Huang CL, Kuo E. Mechanism of hypokalemia in magnesium deficiency. J Am Soc Nephrol. 2007;18(10):2649–52. https://doi.org/10.1681/ASN.2007070792.
27. Wilcox CS, Testani JM, Pitt B. Pathophysiology of Diuretic Resistance and Its Implications for the Management of Chronic Heart Failure. Hypertension. 2020;76(4):1045–54. https://doi.org/10.1161/hypertensionaha.120.15205.
28. Mullens W, Damman K, Harjola VP, Mebazaa A, Brunner-La Rocca HP, Martens P, et al. The use of diuretics in heart failure with congestion - a position statement from the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail. 2019;21(2):137–55. https://doi.org/10.1002/ejhf.1369.
29. Felker GM, Lee KL, Bull DA, Redfield MM, Stevenson LW, Goldsmith SR, et al. Diuretic strategies in patients with acute decompensated heart failure. N Engl J Med. 2011;364(9):797–805. https://doi.org/10.1056/nejmoa1005419.
30. Brisco-Bacik MA, Ter Maaten JM, Houser SR, Vedage NA, Rao V, Ahmad T, et al. Outcomes Associated With a Strategy of Adjuvant Metolazone or High-Dose Loop Diuretics in Acute Decompensated Heart Failure: A Propensity Analysis. J Am Heart Assoc. 2018;7(18):e009149. https://doi.org/10.1161/jaha.118.009149.
31. Trullàs JC, Morales-Rull JL, Casado J, Carrera-Izquierdo M, Sánchez-Marteles M, Conde-Martel A, et al. Combining loop with thiazide diuretics for decompensated heart failure: the CLOROTIC trial. Eur Heart J. 2023;44(5):411–21. https://doi.org/10.1093/eurheartj/ehac689.
32. Wu L, Rodriguez M, El Hachem K, Krittanawong C. Diuretic Treatment in Heart Failure: A Practical Guide for Clinicians. J Clin Med. 2024;13(15):4470. https://doi.org/10.3390/jcm13154470.
33. Weiss JN, Qu Z, Shivkumar K. Electrophysiology of Hypokalemia and Hyperkalemia. Circ Arrhythm Electrophysiol. 2017;10(3):e004667. https://doi.org/10.1161/circep.116.004667.
34. Patel RB, Tannenbaum S, Viana-Tejedor A, Guo J, Im K, Morrow DA, et al. Serum potassium levels, cardiac arrhythmias, and mortality following non-ST-elevation myocardial infarction or unstable angina: insights from MERLIN-TIMI 36. Eur Heart J Acute Cardiovasc Care. 2017;6(1):18–25. https://doi.org/10.1177/2048872615624241.
35. Goyal A, Spertus JA, Gosch K, Venkitachalam L, Jones PG, Van den Berghe G, et al. Serum potassium levels and mortality in acute myocardial infarction. JAMA. 2012;307(2):157–64. https://doi.org/10.1001/jama.2011.1967.

36. Macdonald JE, Struthers AD. What is the optimal serum potassium level in cardiovascular patients? J Am Coll Cardiol. 2004;43(2):155–61. https://doi.org/10.1016/j.jacc.2003.06.021.
37. Aldahl M, Jensen AC, Davidsen L, Eriksen MA, Møller Hansen S, Nielsen BJ, et al. Associations of serum potassium levels with mortality in chronic heart failure patients. Eur Heart J. 2017;38(38):2890–6. https://doi.org/10.1093/eurheartj/ehx460.
38. Núñez J, Bayés-Genís A, Zannad F, Rossignol P, Núñez E, Bodí V, et al. Long-Term Potassium Monitoring and Dynamics in Heart Failure and Risk of Mortality. Circulation. 2018;137(13):1320–30. https://doi.org/10.1161/circulationaha.117.030576.
39. Linde C, Qin L, Bakhai A, Furuland H, Evans M, Ayoubkhani D, et al. Serum potassium and clinical outcomes in heart failure patients: results of risk calculations in 21 334 patients in the UK. ESC Heart Fail. 2019;6(2):280–90. https://doi.org/10.1002/ehf2.12402.
40. Ahmed A, Zannad F, Love TE, Tallaj J, Gheorghiade M, Ekundayo OJ, et al. A propensity-matched study of the association of low serum potassium levels and mortality in chronic heart failure. Eur Heart J. 2007;28(11):1334–43. https://doi.org/10.1093/eurheartj/ehm091.
41. Kardalas E, Paschou SA, Anagnostis P, Muscogiuri G, Siasos G, Vryonidou A. Hypokalemia: a clinical update. Endocr Connect. 2018;7(4):R135–46. https://doi.org/10.1530/ec-18-0109.
42. Laslett DB, Cooper JM, Greenberg RM, Yesenosky GA, Basil A, Gangireddy C, et al. Electrolyte Abnormalities in Patients Presenting With Ventricular Arrhythmia (from the LYTE-VT Study). Am J Cardiol. 2020;129:36–41. https://doi.org/10.1016/j.amjcard.2020.04.051.
43. Riley RD, Debray TPA, Collins GS, Archer L, Ensor J, van Smeden M, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Stat Med. 2021;40(19):4230–51. https://doi.org/10.1002/sim.9025.

Figures

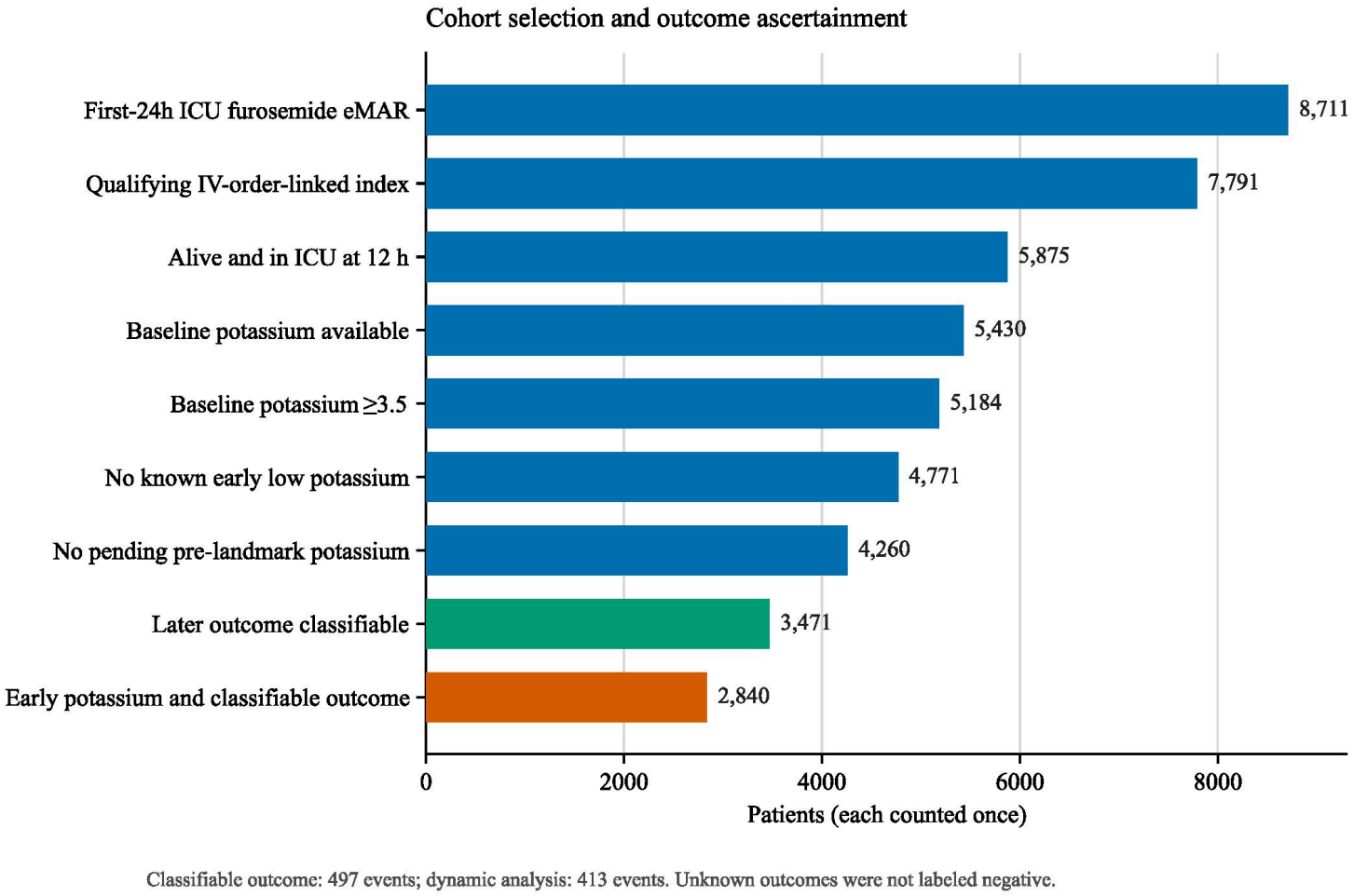


Figure 1

Cohort selection and outcome ascertainment. Patient-level nested subsets lead from 41,625 adults with an admission carrying a cardiovascular diagnosis to the 2,840-person dynamic analysis sample, including 413 recorded hypokalemia events. Among 3,516 risk-set patients with an early potassium result, 676 lacked enough later testing to classify the outcome and were not counted as non-events. Plotting code was developed with OpenAI Codex assistance from aggregate cohort counts.

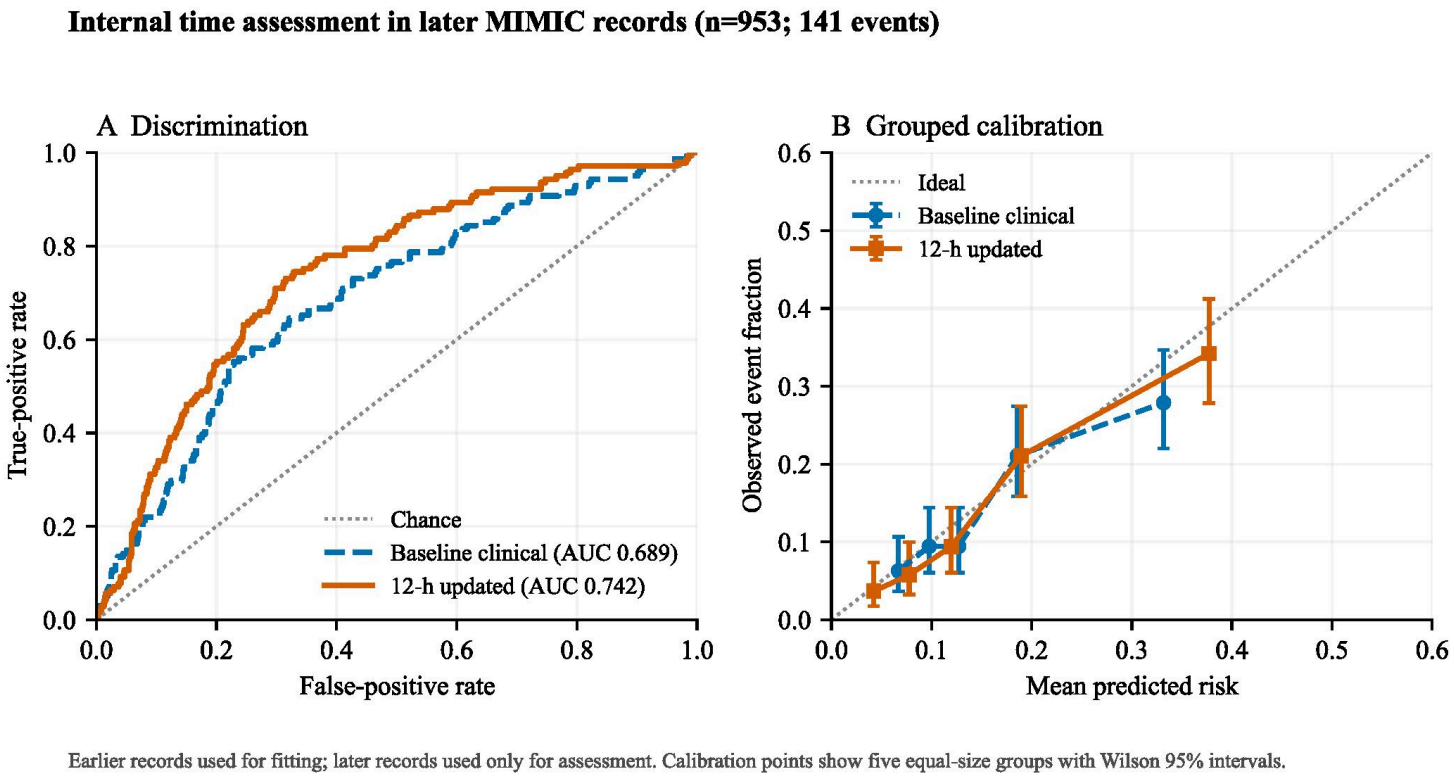


Figure 2

Internal temporal assessment. (A) Paired receiver operating characteristic curves for baseline and updated models in 953 later patients (141 events); the AUC difference was 0.054 (95% patient-bootstrap interval 0.019–0.089; paired DeLong P=0.004). (B) Five-group calibration with Wilson 95% intervals. Models were fitted in earlier same-hospital records without later refitting; calendar-year intervals were inferred from deidentified anchor fields. Plotting code was developed with OpenAI Codex assistance from analysis outputs.

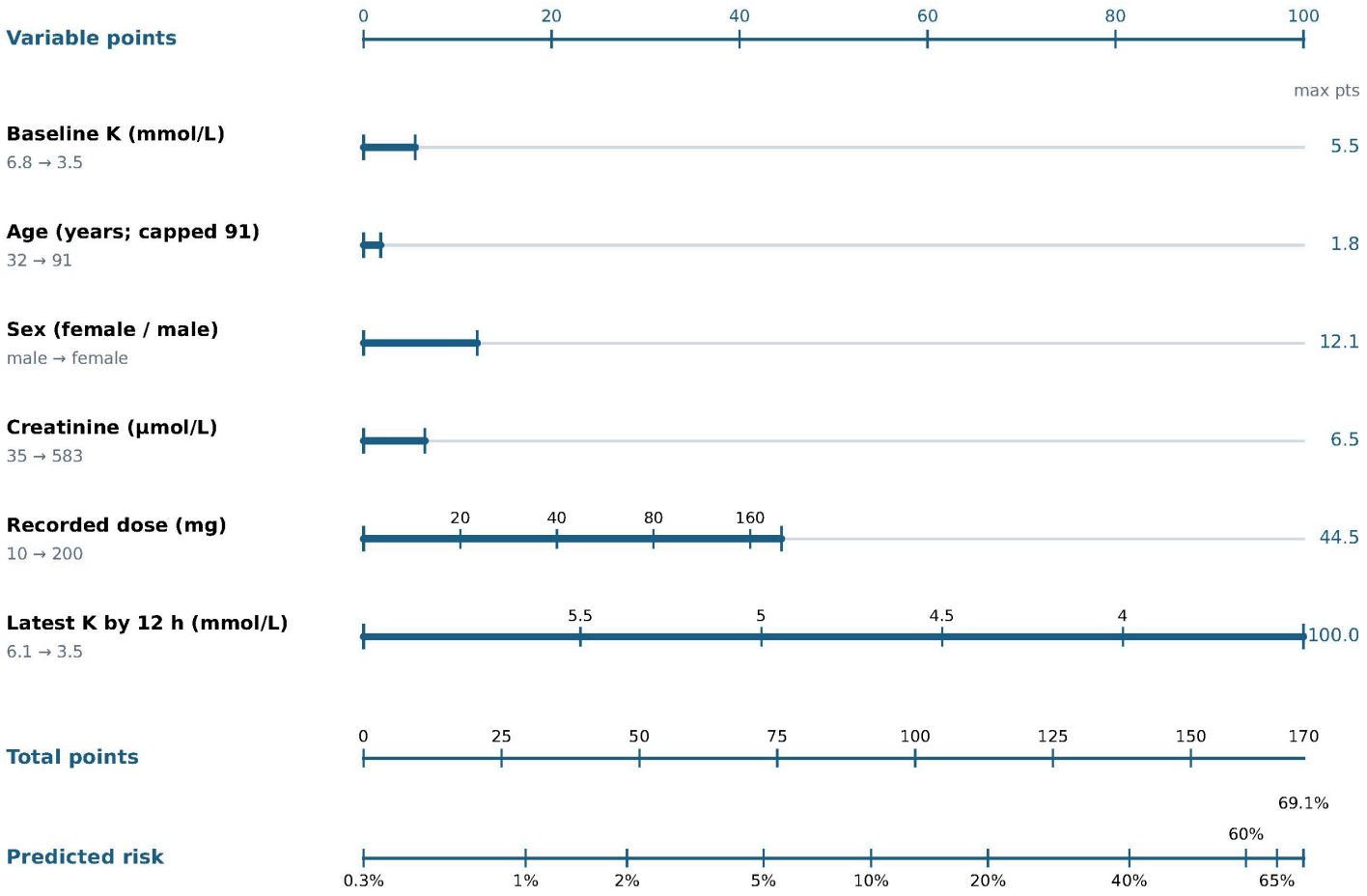


Figure 3

Research-only nomogram of the MIMIC six-variable updated model. Points reproduce the apparent fitted coefficients for baseline potassium, age capped at 91 years, recorded sex (male=1, female=0), log2 baseline creatinine, log2 recorded index furosemide dose, and the latest potassium available by 12 hours. Variable ranges approximate the development-sample 1st–99th percentiles; the displayed total-point range corresponds to approximately 0.3%–69.1% predicted risk. The complete equation, SI-unit input rule, and missing-creatinine rule are in Additional file 1. The graph is for research interpretation, not clinical decision making. Plotting code was developed with OpenAI Codex assistance from fitted coefficients.

Supplementary Files

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

- Additionalfile1Modelspecification.pdf
- Additionalfile2NWICUexploratoryROC.pdf

- Additionalfile3INSPIREexploratoryROC.pdf