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.

Incremental prognostic information from first day archived ST summaries for hospital mortality in a multicentre landmark cohort

Hao Zhu
Nanjing First Hospital
Mingxue Shao
Nanjing First Hospital
Nailiang Tian
Nanjing First Hospital
Xiaobo Li
liber0908@163.com

Nanjing First Hospital

Research Article

Keywords: ST segment monitoring, medical informatics, hospital mortality, landmark analysis, incremental prediction, model evaluation, archived measurements

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11259356/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

Archived monitor measurements can be associated with outcomes without adding useful prognostic information. We evaluated the additional information from simple first-day ST temporal summaries and its sensitivity to channel availability.

Methods

This retrospective eICU study evaluated adults with sufficient first-day ST records who remained in ICU and hospital beyond a 24-hour landmark. One hospitalization per patient was selected without using outcomes. Hospital-blocked five-fold validation compared clinical and Acute Physiology and Chronic Health Evaluation (APACHE) IVa models with mean and operational burden, then slope and root mean square of successive differences (rMSSD). Paired intervals used 1000 hospital resamples of fixed out-of-fold predictions. Exploratory full-refit analyses used 100 hospital-frequency draws per family with original folds retained. A post hoc extension repeated comparisons in patients with a stable valid-channel set.

Results

Among 39020 patients in 127 hospitals, 3435 died. Temporal addition increased clinical area under the receiver operating characteristic curve (AUC) by 0.00034 (conditional 95% confidence interval − 0.00009 to 0.00081). Its APACHE increment was 0.00045 (− 0.00009 to 0.00095; n = 38646). Channel sets changed in 34.73% of patients. Among 25468 stable-channel patients with 2251 deaths, the clinical temporal increment was 0.00089 (0.00011 to 0.00160); the APACHE increment was 0.00066 (0.00015 to 0.00115; n = 25216). Temporal full-refit ranges spanned zero in both model families in both populations, and Brier-score improvement was not established. Joint clinical associations persisted in the original analysis.

Conclusions

Simple temporal ST summaries contributed small discrimination increments, with positive conditional intervals in the stable-channel sensitivity but uncertainty when training was repeated. The findings depend on population, comparator and uncertainty scope; they do not establish equivalence, clinical ischaemia or a validated ST alarm.

Trial registration

Not applicable to this retrospective noninterventional database analysis.

Background

Electrocardiographic monitoring is routine in intensive care, and clinically verified ST changes can inform assessment [1, 2]. The numeric outputs archived by monitoring systems offer a different, widely available record of these measurements. The eICU Collaborative Research Database stores ST1–ST3 in its periodic vital-sign table, generally as five-minute medians of one-minute monitor averages [3, 4]. These records permit investigation of first-day monitoring patterns across hospitals, but their physical units and anatomical lead identities are not established by the source documentation. Prognostic findings from these archived measurements therefore require interpretation distinct from findings based on standard electrocardiograms or adjudicated ischaemia.

Previous studies have associated clinical ST findings with myocardial injury and adverse outcomes in chest-pain, perioperative and intensive-care populations [5–8]. A recent study of 307 non-cardiac intensive-care patients examined admission 12-lead ST deviation and 30-day mortality, including its contribution beyond SOFA [7]. Another study combined first-day eICU ST1–ST3 with other vital-sign series in signal-processing models of ICU mortality [9]. These studies establish relevant clinical and computational context. They also leave a focused question for the present analysis: how much prognostic information do simple temporal ST summaries contribute after recorded ST level and burden and illness severity are already represented?

An adjusted association does not establish improved prediction from a model containing related information [10]. Signed mean and operational absolute burden summarise ST level, whereas the slope over elapsed hours and the root mean square of successive differences (rMSSD) summarise trend and short-term variation. Two nested comparisons distinguish the contribution of level and burden beyond a clinical baseline from the additional contribution of the temporal summaries. Paired evaluation in the same patients and held-out hospitals, including comparison with APACHE IVa risk [11], helps separate these questions. A further measurement concern is that changes in the available channels may alter their composite value and consequently its temporal summaries.

We evaluated the incremental predictive performance of first-day archived ST summaries for subsequent hospital death among adults who remained in intensive care and hospital beyond a 24-hour landmark and had sufficient ST records. The principal comparison added slope and rMSSD to models already containing signed mean and operational burden; clinical and APACHE risk baselines were assessed separately. Jointly adjusted associations were secondary to this incremental-information question. We also conducted a post hoc measurement sensitivity, specified after the initial analyses, to assess these comparisons among patients with a stable valid-channel set. The emphasis on the incremental-information comparison was sharpened after reviewing the original results. The study evaluates the specified archived measurements and feature definitions, rather than an ischaemia diagnostic test or a clinical ST alarm.

Methods

Study design and data source

We used the eICU Collaborative Research Database version 2.0, a deidentified US multicentre database covering 2014–2015 [3, 12, 13]. The 24-hour target population, feature definitions and validation design were retrospective revisions of a broader 48-hour analysis. A dated revision plan preceded the corrected results, and a validation amendment followed preliminary fits. The study was not prospectively preregistered. Reporting follows the relevant elements of RECORD and TRIPOD + AI [14, 15]. Supplementary methods and results are provided in Additional file 1. The channel-stability extension was specified on 4 October 2026 after the initial results were known and before its sensitivity results were computed. Its purpose was to examine measurement robustness; it was not prospectively preregistered.

Risk set and outcome

The landmark was 1440 minutes after ICU admission. The starting ST pool comprised adults aged at least 18 years at their first ICU unit visit within a hospitalization, with ICU stays of at least 24 hours and at least 50 quality-controlled ST observations within 48 hours. We required a known hospital discharge outcome, ICU and hospital discharge offsets strictly greater than 1440 minutes, at least 50 valid observations in minutes 0–1440, finite feature values, positive valid observed duration and at least one exact five-minute observation pair. The first-day count requirement was stricter than the original 48-hour requirement. The outcome was death before hospital discharge after the landmark. Unknown discharge outcomes were excluded.

One eligible hospitalization per patient was selected before the APACHE completeness restriction, using the smallest SHA256 hash of “20261001|patientunitstayid”, with the identifier as a tie-breaker. Selection was independent of outcome and did not necessarily identify the earliest lifetime admission. The primary cohort required a valid APACHE IVa score and recorded Female or Male sex. Six records with unrecorded or other sex, all survivors, were excluded after preliminary fits showed complete separation; final models and resampling analyses were rerun. Table 1 retains the 39026-person stage before this amendment.

Table 1
Cohort flow

Stage	Hospital stays	Patients	Hospitals	Deaths
Supplied first unit visit ST pool	50033	44659	142	4073
Known hospital discharge outcome	49480	44206	141	4073
ICU and hospital exits after 24 hours	49412	44150	141	4065
At least 50 valid first day ST records	48381	43258	139	3970
All four ST summaries computable	48381	43258	139	3970
One eligible hospitalization per patient	43258	43258	138	3706
Valid APACHE IVa score	39026	39026	127	3435
Recorded Female or Male sex	39020	39020	127	3435
Common valid APACHE IVa risk subset	38646	38646	127	3400

Sequential restrictions are shown, followed by the common-risk subset. Deaths before the landmark exit restriction are source-recorded hospital deaths. The six-record sex exclusion followed preliminary fits; final analyses used 39020 patients. Initial ICU and hospital exit exclusions overlap.

Archived ST measurements and summaries

Channel values within the operational archived-scale range − 5 to 5 were retained. At each timestamp, the representative ST value was the median of three valid channels, the average of two, or the single available value. Repeated offsets were merged by their median. Values were restricted to minutes 0–1440 and to the period before recorded ICU or hospital exit. Interchannel spread greater than 2 was flagged but retained. We report values in archived units: the range limit, spread flag and burden threshold of 0.2 are processing conventions, not clinical ischaemia thresholds. The source periodic values were not bedside-validated; range filtering did not adjudicate physiological validity or ischaemia [4].

The signed mean used trapezoidal endpoint weights for consecutive observations separated by more than zero and no more than 30 minutes, divided by the total included duration. Operational burden applied the same weights to the indicator |archived ST| ≥ 0.2. Longer gaps were omitted, with no extrapolation after the final observation. Slope was the ordinary least-squares coefficient of ST against elapsed hours. The root mean square of successive differences (rMSSD) used adjacent observations exactly five minutes apart. Supplementary Methods and Tables S2 and S10 give the formulas and coverage distributions.

Clinical predictors and recorded timing

The clinical base included age, recorded sex, APACHE IVa score, ICU type, ST count and valid duration, qualifying troponin count and tier, recorded history and comorbidities, acute myocardial infarction, ST-elevation myocardial infarction, sepsis, cardiac arrest, and early circulatory-support and ventilation-initiation indicators. All covariates and the complete 38-parameter joint model are listed in Table S6. Disease and treatment indicators represent documentation within the specified window.

Troponin samples had to occur in minutes 0–1440, with revised result entry between sampling and the landmark; negative values and entry-before-sample inconsistencies were excluded [16]. Units were harmonized using a nonempty system-unit field before the interface field. Qualifying peaks were grouped as no record, below 0.04, 0.04 to below 0.12, 0.12 to below 0.4, or at least 0.4 ng/mL. These adjustment tiers combined troponin I and T assays and were not assay-specific myocardial-injury definitions [17–19]. Twenty-one records in 11 primary patients had discordant system ng/mL and interface mg/mL labels; the primary pipeline used the system value. A point-only sensitivity assigned these results to an unknown-unit category while retaining patients, test counts, ST features and folds (Table S13).

History, diagnosis, treatment, infusion and respiratory-care records were restricted by their available offsets to the landmark, as detailed in Supplementary Methods. Ventilation denoted a documented initiation in minutes 0–1440 and did not fully capture pre-ICU ventilation. APACHE IVa was selected explicitly, with no duplicate IVa rows in the selected cohort. Negative score and risk sentinels were treated as missing, and valid risks were restricted to [0,1] [20]. These rules define retrospective first-day inputs; the archive does not establish the exact bedside availability of every predictor, including the APACHE calculation time.

Association analysis and missingness

Logistic regression included all four ST summaries and the clinical base. Continuous effects were scaled by the cohort sample standard deviation (SD) and entered as linear logit terms. We report adjusted odds ratios (ORs), 95% confidence intervals (CIs) and two-sided P values using hospital-clustered robust variance. The four ST terms were also compared with a Bonferroni reference of 0.0125. Hospital fixed-effect sensitivity excluded hospitals without outcome variation; an all-hospitalization sensitivity used two-way patient- and hospital-clustered variance.

The primary analysis required complete APACHE scores and recorded Female or Male sex. A prediction sensitivity included missing scores using training-fold median imputation and a missing-score indicator, with the same sex requirement. No sex values were imputed, and multiple imputation was not performed.

Nested models and held out hospital evaluation

M0 was the clinical base. M2 added mean and burden; M3 retained those terms and added slope and rMSSD. M1 added only the first ST value to M0 as a separate comparator. Models M0, M1, M2 and M3 had 34, 35, 36 and 38 parameters including the intercept. The key temporal comparison was M3 versus M2; M3 versus M0 evaluated all four summaries together. No a priori sample-size calculation was performed; the primary joint model had 3435 deaths for 38 parameters. Paired comparisons used identical patients and hospital folds.

Hospital-blocked five-fold cross-validation used non-shuffled GroupKFold allocation, with each hospital assigned to one test fold. One-patient selection prevented patient overlap. In the all-hospitalization sensitivity, other records of test patients were removed from training. Scaling, categorical encoding, imputation and recalibration were fitted within training folds. We report pooled out-of-fold area under the receiver operating characteristic curve (AUC), Brier score and jointly estimated calibration intercept and slope [10]. Calibration CIs used hospital-clustered robust variance. Mean fold AUC and fold counts are provided separately in Tables S4, S8, S9 and S14.

Paired 95% percentile CIs used 1000 hospital-cluster resamples of fixed out-of-fold predictions (seed 20261001). These intervals are conditional on the fitted pipelines. An exploratory hospital bootstrap refitted the training transformations and three nested models on identical hospital-frequency draws, separately for the clinical and APACHE families. Each family used 100 draws and 15 fits per draw, retaining the original hospital-to-fold mapping. We report its 95% percentile ranges as a sensitivity to training and hospital sampling; the small number of draws limits tail precision.

APACHE risk comparators and sensitivities

On patients with valid APACHE IVa risk, we compared raw risk, training-fold logistic recalibration of its logit, recalibration plus mean and burden, and recalibration plus all four summaries. These models did not add ICU type or troponin terms. Clinical models were refitted on the same patients, allowing direct comparison with the stronger risk baseline.

Prediction sensitivities required at least 12 hours of valid ST duration, excluded eight hospitals flagged by an earlier scale audit, included missing APACHE scores, used all eligible hospitalizations with test-patient purging, or added troponin samples from the preceding 24 hours under the entry-time restriction. The eight-hospital list was fixed from the preceding audit rather than rederived in the final cohort (Supplementary Methods). A discharge-year split trained on 2014 and tested on 2015; it separated patients but allowed shared hospitals. Year refers to hospital discharge, not admission [21].

Exploratory decision curves evaluated hypothetical mortality-risk thresholds of 5–20% for reassessment. Net benefit was expressed as net true-positive equivalents per 1000 patients [22]. Analyses used Python, DuckDB, pandas, NumPy, scikit-learn and statsmodels; software versions and reproducibility information are supplied with the supplementary material.

Channel availability sensitivity

For the channel-stability sensitivity, the valid-channel set at each retained offset was the union of channels with values within − 5 to 5 across duplicate records. Patients were classified as stable if this set did not change between retained offsets during minutes 0–1440. No outcome information was used for this classification, and all original primary eligibility and hospitalization-selection rules were retained. Clinical models used the original primary hospital-to-fold map; common-risk comparisons used the original common-risk map. Filtering did not trigger a new GroupKFold allocation. Transformations and models were refitted within the restricted training folds. Fixed-prediction intervals used 1000 hospital resamples (seed 20261004). The stable primary clinical family and stable common-risk APACHE family each underwent 100 complete frequency-weighted refits (seeds 61320261004 and 61420261004), retaining the original folds. Stable common-risk clinical comparisons had conditional intervals only. No failed draws were discarded or replenished. This restriction changes the target population and does not validate physical units, anatomical leads or ischaemia.

Results

Cohort and recording coverage

The starting pool contained 50033 hospitalizations from 44659 patients in 142 hospitals. After outcome, exit-time and ST-coverage restrictions, 48381 hospitalizations were eligible. One-hospitalization selection retained 43258 patients. The final primary cohort comprised 39020 patients, 3435 hospital deaths (8.803%) and 127 hospitals after APACHE-score and recorded-sex restrictions. The common APACHE-risk subset included 38646 patients and 3400 deaths (Table 1; Fig. 1).

The primary cohort contributed 10126736 valid ST observations. Median recording count was 280 (IQR 259–287), and median valid duration was 1410 minutes (1305–1430). Of 10087716 adjacent intervals, 9982956 (98.962%) were exactly five minutes and 22774 (0.226%) exceeded 30 minutes. At least one long gap occurred in 12275 patients. The spread flag affected 313136 timestamps (3.092%) in 6914 patients. Qualifying first-day troponin records were present in 10320 patients (26.45%). Tables 2, S2, S10 and S15 describe patient characteristics, recording coverage and channel distributions.

Table 2
Primary cohort and feature distributions **Panel A Continuous distributions**

Characteristic	Median and IQR	Mean and SD
Age years	66 (54 to 77)	64 (17)
APACHE IVa score	54 (40 to 71)	58 (26)
ST observations	280 (259 to 287)	260 (49)
Valid ST duration, hours	23.50 (21.75 to 23.83)	21.82 (3.82)
Time weighted mean	0.009470 (-0.204179 to 0.221345)	0.010294 (0.495225)
Operational burden	0.573256 (0.168473 to 0.937063)	0.542615 (0.373300)
Slope per hour	0.000000 (-0.006506 to 0.006854)	0.000339 (0.025072)
Exact 5 minute rMSSD	0.076020 (0.057685 to 0.106962)	0.101412 (0.116222)
Valid coverage fraction	0.979167 (0.906250 to 0.993056)	0.909194 (0.159054)
Panel B Categorical and documentation summaries		

Clinical incremental performance

Pooled out-of-fold AUC was 0.811968 for M0, 0.812084 for M1, 0.813885 for M2 and 0.814223 for M3. Adding mean and burden increased AUC by 0.001918 (conditional 95% CI 0.000395–0.003336). Adding slope and rMSSD after those terms changed AUC by 0.000337 (− 0.000093–0.000808) and Brier score by − 0.000003 (− 0.000055–0.000046). All four summaries versus the clinical base produced an AUC difference of 0.002255 (0.000658–0.003729), while its Brier interval crossed zero. M3 calibration intercept and slope were − 0.0313 and 0.9758 (Tables 3 and 4; Figs. 2 and S1).

Table 3
Held out hospital performance in primary and common risk cohorts

Model	Pooled AUC	Brier	Calibration intercept	Calibration slope
Primary cohort Clinical models: N = 39020; deaths = 3435				
M0 clinical base	0.811968	0.067412	-0.0404	0.9731
M1 plus first ST	0.812084	0.067410	-0.0395	0.9728
M2 plus mean and burden	0.813885	0.067267	-0.0303	0.9762
M3 plus slope and rMSSD	0.814223	0.067264	-0.0313	0.9758
Common valid-risk subset Clinical models: N = 38646; deaths = 3400				
M0 clinical base	0.811529	0.067437	-0.0516	0.9710
M2 plus mean and burden	0.813363	0.067282	-0.0436	0.9744
M3 plus slope and rMSSD	0.813811	0.067266	-0.0431	0.9746
Common valid-risk subset APACHE risk models: N = 38646; deaths = 3400				
Raw APACHE IVa risk	0.838385	0.069209	-0.6737	0.8993
Recalibrated APACHE risk	0.837890	0.065501	-0.0027	0.9985
APACHE plus mean and burden	0.838622	0.065432	-0.0023	0.9985
APACHE plus all four ST summaries	0.839071	0.065406	-0.0033	0.9982

Primary clinical block: N = 39020 and 3435 deaths. Common-subset clinical and APACHE risk blocks: N = 38646 and 3400 deaths. Raw risk is not refitted; all recalibration and augmentation are trained within folds. Calibration intercept and slope are estimated jointly. Mean fold AUC and coefficient intervals are in Table S4.

Table 4
Paired conditional incremental performance

Population	Comparison	AUC change and 95% CI	Brier change and 95% CI
Primary	First ST versus base	0.000117 (-0.000437 to 0.000564)	-0.000002 (-0.000045 to 0.000043)
Primary	Mean and burden versus base	0.001918 (0.000395 to 0.003336)	-0.000145 (-0.000285 to 0.000002)
Primary	Temporal addition versus mean and burden	0.000337 (-0.000093 to 0.000808)	-0.000003 (-0.000055 to 0.000046)
Primary	All four versus base	0.002255 (0.000658 to 0.003729)	-0.000148 (-0.000300 to 0.000020)
Common APACHE risk	Mean and burden versus recalibrated risk	0.000732 (-0.000124 to 0.001467)	-0.000069 (-0.000181 to 0.000053)
Common APACHE risk	Temporal addition versus mean and burden	0.000449 (-0.000086 to 0.000949)	-0.000026 (-0.000087 to 0.000032)
Common APACHE risk	All four versus recalibrated risk	0.001181 (0.000141 to 0.002143)	-0.000095 (-0.000224 to 0.000051)

Extended minus reference. Positive AUC change and negative Brier change favour extension. Intervals use 1000 hospital resamples of fixed out of fold predictions, conditional on the fitted pipelines. Separate 100 draw full refit sensitivity ranges are given in Table S11. All 19 reported paired comparisons are in Table S5.

The clinical full-refit bootstrap gave a median temporal AUC increment of 0.000233 (95% percentile range − 0.000259 to 0.000859) and Brier difference of 0.000011 (− 0.000073 to 0.000080). Mean and burden versus M0 had a median AUC increment of 0.001716 (0.000145–0.004205); all four versus M0 had a median increment of 0.002146 (0.000236–0.004520; Table S11).

Common APACHE risk comparisons

On the common 38646-patient subset, raw APACHE IVa risk had AUC 0.838385 and Brier score 0.069209. Training-fold recalibration yielded AUC 0.837890 and Brier score 0.065501. Adding mean and burden changed AUC by 0.000732 (conditional 95% CI − 0.000124–0.001467); further temporal addition changed it by 0.000449 (− 0.000086–0.000949). All four versus recalibrated APACHE risk yielded an AUC difference of 0.001181 (0.000141–0.002143) and Brier difference of − 0.000095 (− 0.000224–0.000051). The full clinical model had a lower AUC of 0.813811 on these same patients (Tables 3 and 4).

The APACHE full-refit bootstrap gave a median temporal AUC difference of 0.000340 (95% percentile range − 0.000167–0.001060) and an all-four versus base difference of 0.000935 (− 0.000244–0.002712). Both Brier ranges crossed zero. The common-subset clinical temporal increment was 0.000448 (conditional 95% CI 0.000046–0.000864), showing that the small increments varied with population and comparator (Tables S5 and S11).

Channel availability and incremental information

The valid-channel set changed in 13552 patients (34.73%), leaving 25468 stable-channel patients with 2251 deaths in 121 hospitals. Stable and switching groups had similar descriptive mortality (8.84% and 8.74%), while median ST counts were 282 and 277 and median rMSSD was 0.0724 and 0.0856 archived units, respectively. The stable group contained no retained duplicate offset with discordant valid-channel sets. These comparisons are descriptive and do not identify effects of changing channel availability (Table S16).

Within the stable primary population, temporal addition changed AUC by 0.000890 (0.000106 to 0.001599) and Brier by 0.000002 (-0.000073 to 0.000075). The clinical full-refit median AUC increment was 0.000674 (-0.000405 to 0.002079). Among 25216 stable patients with valid APACHE risk (2228 deaths), temporal addition beyond recalibrated APACHE plus mean and burden changed AUC by 0.000659 (0.000149 to 0.001152) and Brier by 0.000023 (-0.000054 to 0.000115). Its full-refit median AUC increment was 0.000508 (-0.000129 to 0.001534). Both temporal full-refit ranges and all corresponding temporal Brier intervals or ranges included zero (Table 5; Fig. 3; Tables S17–S20).

Table 5
Temporal increments in stable channel populations

Population and family	N	Observed AUC change and conditional 95% CI	Full refit AUC median and 95% range	Observed Brier change and conditional 95% CI
Stable primary clinical	25468	0.000890 (0.000106 to 0.001599)	0.000674 (-0.000405 to 0.002079)	0.000002 (-0.000073 to 0.000075)
Stable common APACHE risk	25216	0.000659 (0.000149 to 0.001152)	0.000508 (-0.000129 to 0.001534)	0.000023 (-0.000054 to 0.000115)

All changes are extended minus reference: slope and rMSSD added to mean and burden. Positive AUC and negative Brier favour extension. Conditional intervals use 1000 hospital resamples of fixed OOF predictions; full-refit ranges use 100 complete training-refit draws and have limited tail precision. Both comparisons are paired within their own stable population. Full model results and Brier ranges are in Tables S17–S19.

In that stable common-risk subset, all four summaries versus recalibrated APACHE had a full-refit median AUC increment of 0.001293 (0.000350 to 0.002964), while its Brier range included zero. The common-subset clinical temporal increment was 0.000973 (conditional 95% CI 0.000269 to 0.001644); this clinical subset did not undergo a separate full-refit analysis. These results show small increments whose interpretation depends on the comparator and uncertainty scope. Different-population AUC changes are not paired estimates of a channel-switching effect.

Secondary joint adjusted associations

In the jointly adjusted model, higher mean and slope were associated with lower odds of hospital death, while higher operational burden was associated with higher odds. Per-SD ORs were 0.9103 (95% CI 0.8767–0.9452; P = 9.919×10^{-7}) for mean, 1.0817 (1.0319–1.1339; P = 0.001089) for burden and 0.9536 (0.9222–0.9861; P = 0.005409) for slope. All three met the four-feature Bonferroni reference of 0.0125. The rMSSD association was uncertain (OR 1.0304, 0.9847–1.0782; P = 0.1961; Table 6; Fig. 4).

Table 6
Joint associations with subsequent hospital death

ST summary	Cohort SD	Adjusted OR	95% CI	P
Time weighted mean	0.49522452	0.9103	0.8767 to 0.9452	9.92e-07
Operational burden	0.37330043	1.0817	1.0319 to 1.1339	0.0011
Slope per hour	0.02507190	0.9536	0.9222 to 0.9861	0.0054
Exact 5 minute rMSSD	0.11622227	1.0304	0.9847 to 1.0782	0.1961

N = 39020; 3435 deaths; 127 hospitals. All four summaries were mutually adjusted with the clinical terms in Table S6, 38 parameters including intercept. Intervals use hospital clustered robust variance. Slope SD is in archived units per hour; other feature scales are as described in Table 2.

Mean, burden and slope retained their directions in the hospital fixed-effect and all-hospitalization analyses. The fixed-effect model included 38932 patients, 3434 deaths and 109 hospitals after excluding 18 hospitals without outcome variation. The all-hospitalization model included 43706 hospitalizations from 39283 patients and 3678 deaths. rMSSD remained uncertain in both analyses (Table S3).

Other sensitivity findings

Coverage, flagged-hospital exclusion, missing-score, repeated-hospitalization, troponin-window, discharge-year and point-only unit sensitivities are reported in Additional file 1. Small positive conditional temporal increments occurred in some specifications. Exploratory decision curves use hypothetical reassessment thresholds and do not establish clinical benefit (Tables S4, S5, S11 and S13; Figure S2).

Discussion

In 39020 adults from 127 hospitals who remained in intensive care and hospital beyond 24 hours, simple temporal ST summaries contributed small increments in discrimination after recorded mean and operational burden were included. In the primary clinical comparison, adding slope and rMSSD increased out-of-fold AUC from 0.813885 to 0.814223: a difference of 0.000337, with a conditional 95% interval from − 0.000093 to 0.000808. The clinical full-refit analysis gave a similarly small median increment, with a percentile range including zero. Mean, burden and slope nevertheless retained conditional associations with hospital death in the joint model. These results illustrate why the association of a monitoring

feature with outcome and its incremental contribution to prediction need separate evaluation. The observed pattern supports limited additional information from the two temporal summaries under the tested definitions and models; it does not establish equivalence or the absence of useful information in other ST representations.

The comparator and population materially affected interpretation. On the same common-risk subset, recalibrated APACHE IVa risk had greater discrimination than the full clinical ST model. Adding all four ST summaries to this risk baseline produced a small positive conditional AUC increment, while its full-refit range and Brier interval included zero. The common-subset clinical temporal interval was above zero, unlike the primary clinical and APACHE temporal intervals. These differences should be retained rather than compressed into a universal negative finding. Prior clinical ECG studies address interpretable ST abnormalities and different outcomes [5–8], whereas the eICU signal-processing study evaluated combinations of several vital-sign series in selected diagnostic cohorts [9]. Our focused comparison concerns the additional value of two ST summaries beyond level, burden and a severity comparator. Differences in measurements, eligibility, outcomes and validation prevent direct comparisons of reported AUCs across these studies.

The measurement construction is central to this interpretation. The signed composite uses available valid channels, whose polarity, anatomical identity and physical units are not verified; combining channels can also obscure opposing deviations. A change in channel availability can change the composite without a corresponding physiological change. In the original primary cohort, 13552 patients (34.73%) had a change in their recorded valid-channel set between retained offsets under the union rule for duplicate offsets. The stable-channel sensitivity produced positive conditional temporal AUC intervals in both the primary clinical and APACHE families, with increments of 0.000890 and 0.000659. Their full-refit ranges nevertheless included zero, and Brier improvement was not established. All four ST summaries beyond recalibrated APACHE had a positive full-refit AUC range in the stable subset, with an uncertain Brier difference. This strengthens the evidence for a small conditional discrimination increment in the restricted population, while preserving uncertainty across training realizations. The differences from the original results cannot be attributed solely to measurement construction because channel restriction also changes the population. Stable channels do not establish physiological validity. Neither channel stability nor statistically supported conditional associations establish ischaemia or a protective effect of a higher signed mean or trend.

Monitoring and laboratory documentation may also reflect clinical indication. Adjusting for recording intensity and documented treatment only partly addresses these processes; treatment can reflect both severity and response to deterioration. Differential clinical contact is a recognized source of ascertainment bias in electronic records [23]. Troponin adjustment here combined assays, assumed numeric tiers and a small set of unit-discordant records. The unknown-unit sensitivity changed aggregate metrics little but showed larger changes for some individual predictions. It did not verify which unit field was correct. Absence of a qualifying troponin, disease or treatment record likewise should not be read as clinical absence [17–19].

Several boundaries remain. The landmark and coverage criteria define a selected population and exclude early deaths and discharges; the findings do not describe admission-time prediction in unselected ICU patients. The primary complete-score and recorded-sex restrictions, outcome-independent hospitalization selection, and monitoring availability further limit transportability. First-day offsets do not verify real-time bedside availability, particularly for APACHE. Clinical diagnoses, treatments and troponin tiers describe documentation and do not fully capture their indication or timing; the unit-discordance sensitivity did not determine the true assay units. Hospital-held-out evaluation remains internal to one historical database, and the discharge-year check allowed shared hospitals. Pooled performance does not establish calibration at each hospital. Conditional prediction intervals omit training variation, while the 100-draw full-refit ranges retain the original folds and have limited tail precision. Retrospective changes to the analysis plan and the new channel-stability analysis require explicit reporting. Continuous terms were modelled linearly, so these results do not exclude additional value from alternative functional forms. The current computational verification supports the reported calculations within its documented scope, including primary-cohort first-day raw-channel checks; it does not reconstruct all historical analysis steps or establish clinical measurement validity. In the repeated-hospitalization sensitivity, resampling hospitals did not capture dependence from repeat patients recorded across hospitals.

The useful contribution is a reproducible benchmark for whether ST temporal summaries add information once level, burden and severity are already known. Future work should use measurements with interpretable units and lead metadata and assess incremental performance against the same strong comparator in independent populations. More complex representations warrant evaluation when they answer this defined question, with feature construction and tuning confined to training data and complete reporting of results. Clinical deployment would additionally require verified predictor availability and a defined action, benefit and cost at a specified prediction time. The exploratory decision curves here used hypothetical reassessment thresholds and do not demonstrate benefit from an ST-triggered intervention [22]. Current findings support cautious study of archived monitoring records, without establishing a replacement for clinical ECG assessment or a validated bedside alarm.

Conclusions

In this selected 24-hour landmark population, simple temporal archived ST summaries added small discrimination increments beyond mean, operational burden and severity. Positive conditional intervals in the stable-channel sensitivity were accompanied by temporal full-refit ranges spanning zero and uncertain Brier improvement. The findings support population- and model-specific incremental information, without establishing equivalence, clinical measurement validity or intervention benefit.

Abbreviations

APACHE
Acute Physiology and Chronic Health Evaluation
AUC

area under the receiver operating characteristic curve
CI
confidence interval
ECG
electrocardiogram
eICU
electronic intensive care unit
ICU
intensive care unit
OR
odds ratio
RECORD
REporting of studies Conducted using Observational Routinely-collected health Data
rMSSD
root mean square of successive differences
SD
standard deviation
ST
electrocardiographic ST segment
TRIPOD
Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.

Declarations

Ethics approval and consent to participate

This study was a retrospective secondary analysis of the deidentified eICU Collaborative Research Database version 2.0, released under the Health Insurance Portability and Accountability Act Safe Harbor framework [12]. Only deidentified records were analysed; participants were not contacted and no intervention was performed. Additional ethics approval and informed consent were not required for this secondary analysis of the deidentified database. The authors confirmed authorized access to the resource.

Consent for publication

Not applicable. The manuscript reports aggregate results and contains no identifiable individual participant material.

Competing interests

The authors declare that they have no competing interests.

Additional files

Additional file 1 (.docx). Title: Supplementary methods and results. Description: Detailed methods, Tables S1–S20 and Figures S1–S2 for the original analysis and post hoc channel-stability extension.

Additional file 2 (.zip). Title: Native aggregate results and figure sources. Description: Machine-readable original and channel-stability aggregates, complete bootstrap-draw outputs, plotting sources, a publication-number crosswalk, indexes and checksums.

Additional file 3 (.zip). Title: Analysis code and local replay instructions. Description: Frozen code snapshots, input schemas, environment requirements, the corrected replay entry, and the channel-stability extension with its dated post hoc plan.

Funding

The authors received no funding for this study.

Author Contribution

Contributions reported by the authors are: HZ, conceptualisation, data curation, formal analysis, methodology and writing the original draft; MS, data curation, validation and review and editing; NT, investigation, visualisation and review and editing; XL, conceptualisation, supervision, project administration and review and editing. All authors read and approved the manuscript.

Acknowledgements

We acknowledge the eICU Collaborative Research Database contributors for making the deidentified resource available for research.

Availability of data and materials

The source dataset is available through PhysioNet as the eICU Collaborative Research Database version 2.0 (https://doi.org/10.13026/C2WM1R) [12]. Access is restricted to credentialed users who meet the required training and data-use conditions; researchers must obtain their own authorized access. The aggregate results supporting the conclusions are included within this article and Additional files 1 and 2. Additional file 3 contains the analysis-code snapshots, input schemas, environment requirements and local replay entries. Patient-level records, individual predictions and fold-membership files are not redistributed.

References

1. Sandau KE, Funk M, Auerbach A, et al. Update to practice standards for electrocardiographic monitoring in hospital settings: a scientific statement from the American Heart Association. Circulation. 2017;136(19):e273–344. 10.1161/CIR.0000000000000527.
2. Fox L, Kirkendall C, Craney M. Continuous ST-segment monitoring in the intensive care unit. Crit Care Nurse. 2010;30(5):33–43. 10.4037/ccn2010823. quiz 44.
3. 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.
4. eICU Collaborative Research Database. vitalPeriodic documentation. https://eicu.mit.edu/eicutables/vitalperiodic/. Accessed 3 October 2026.
5. Jernberg T, Lindahl B, Wallentin L. ST-segment monitoring with continuous 12-lead ECG improves early risk stratification in patients with chest pain and ECG nondiagnostic of acute myocardial infarction. J Am Coll Cardiol. 1999;34(5):1413–9. 10.1016/s0735-1097(99)00370-8.
6. Landesberg G, Mosseri M, Wolf Y, Vesselov Y, Weissman C. Perioperative myocardial ischemia and infarction: identification by continuous 12-lead electrocardiogram with online ST-segment monitoring. Anesthesiology. 2002;96(2):264–70. 10.1097/00000542-200202000-00007.
7. Świstek R, Dadański E, Kurzyca A, et al. Association between ST-segment deviation in electrocardiography and 30-day mortality in non-cardiac critically ill patients: a retrospective single-center study. J Clin Med. 2025;14(14):4911. 10.3390/jcm14144911.
8. Landesberg G, Vesselov Y, Einav S, Goodman S, Sprung CL, Weissman C. Myocardial ischemia, cardiac troponin, and long-term survival of high-cardiac-risk critically ill intensive care unit patients. Crit Care Med. 2005;33(6):1281–7. 10.1097/01.ccm.0000166607.22550.87.
9. Wang S, Jiang Y, Li Q, Zhang W. Intensive Care Unit Patient Outcome Prediction Using v-Support Vector Classification and Stochastic Signal Processing-Based Feature Extraction Techniques: Algorithm Development and Validation Study. JMIR AI. 2025;4:e72671. 10.2196/72671.
10. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128–38. 10.1097/EDE.0b013e3181c30fb2.
11. 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(5):1297–310. 10.1097/01.CCM.0000215112.84523.F0.
12. Pollard T, Johnson A, Raffa J, Celi LA, Badawi O, Mark R. eICU Collaborative Research Database version 2.0. PhysioNet. 2019. https://doi.org/10.13026/C2WM1R
13. Pollard T, Moody BE, Lehman LWH, Gow BJ, Fernandes C, Xie C, et al. PhysioNet as a global platform for biomedical research. Nat Health. 2026;1:792–5. 10.1038/s44360-026-00096-z.

14. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12(10):e1001885. 10.1371/journal.pmed.1001885.
15. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. 10.1136/bmj-2023-078378.
16. eICU Collaborative Research Database. lab documentation. https://eicu.mit.edu/eicutables/lab/. Accessed 3 October 2026.
17. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). Eur Heart J. 2019;40(3):237–69. 10.1093/eurheartj/ehy462.
18. Lim W, Holinski P, Devereaux PJ, et al. Detecting myocardial infarction in critical illness using screening troponin measurements and ECG recordings. Crit Care. 2008;12(2):R36. 10.1186/cc6815.
19. Lim W, Whitlock R, Khera V, et al. Etiology of troponin elevation in critically ill patients. J Crit Care. 2010;25(2):322–8. 10.1016/j.jcrc.2009.07.002.
20. eICU Collaborative Research Database. apachePatientResult documentation. https://eicu.mit.edu/eicutables/apachepatientresult/. Accessed 3 October 2026.
21. eICU Collaborative Research Database. patient documentation. https://eicu.mit.edu/eicutables/patient/. Accessed 3 October 2026.
22. Vickers AJ, van Calster B, Wynants L, Steyerberg EW. Decision curve analysis: confidence intervals and hypothesis testing for net benefit. Diagn Progn Res. 2023;7:11. 10.1186/s41512-023-00148-y.
23. Wang J, Choi M, Buto P, et al. Detection bias in EHR-based research on clinical exposures and dementia. JAMA Netw Open. 2025;8(4):e256637. 10.1001/jamanetworkopen.2025.6637.

Figures

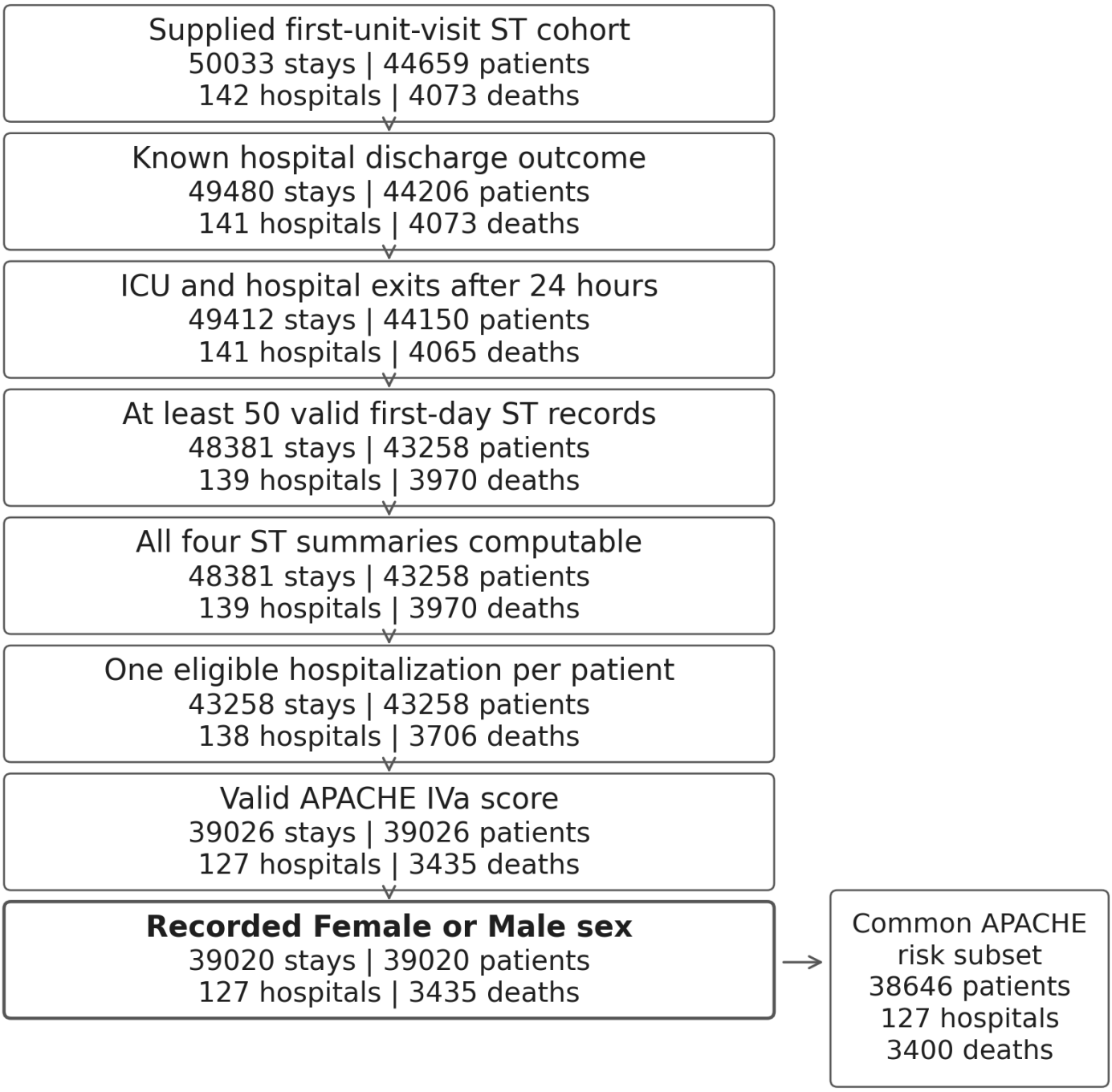


Figure 1

Cohort construction and common risk subset

The supplied ST cohort was restricted sequentially to 39020 patients, 3435 hospital deaths and 127 hospitals. The common valid-APACHE-risk subset contained 38646 patients and 3400 deaths. One hospitalization per patient was selected before APACHE-score and recorded-sex restrictions. Counts before the landmark exit restriction include source-recorded deaths.

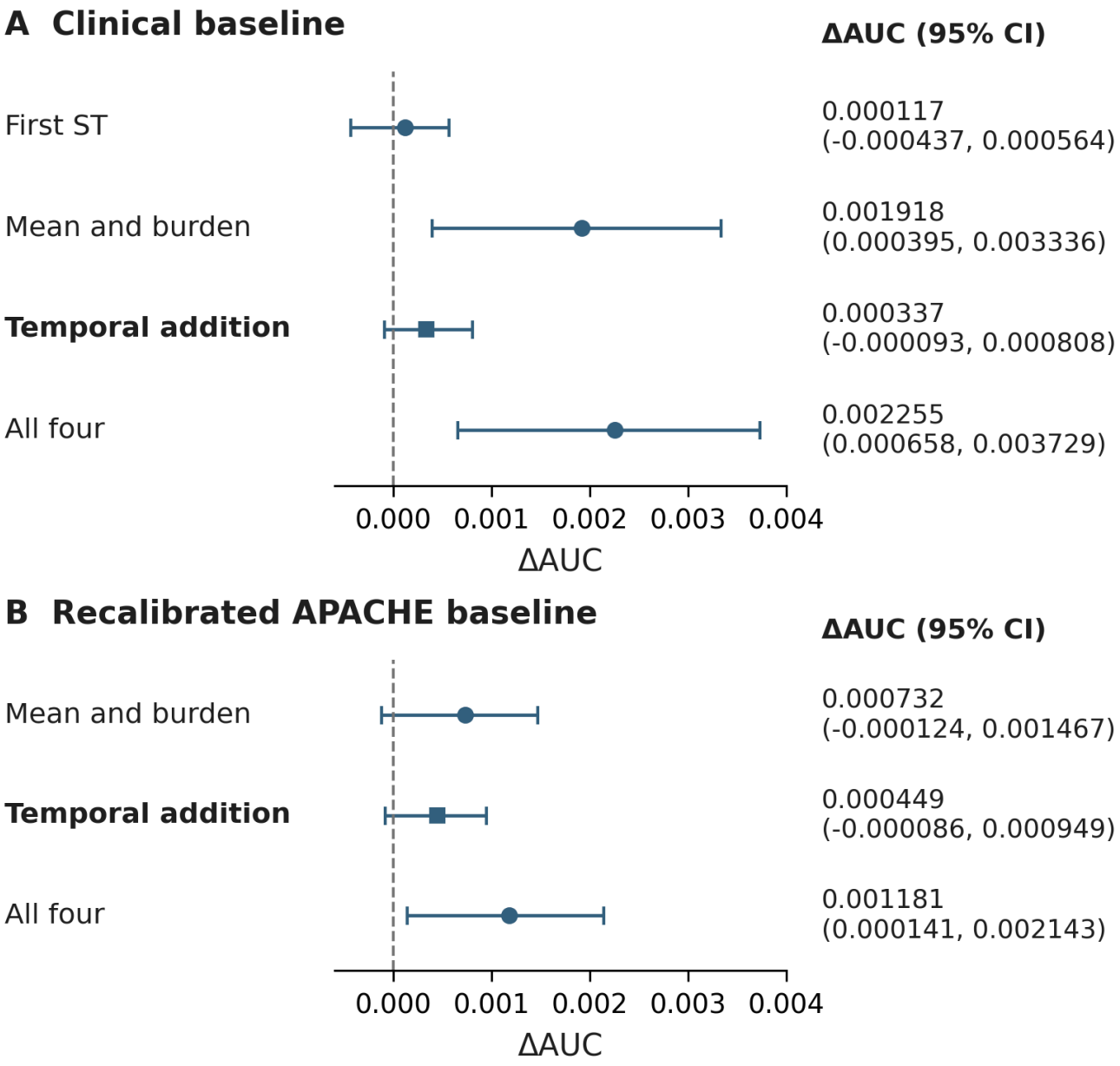


Figure 2

Paired incremental discrimination

Changes in pooled out-of-fold AUC with conditional 95% intervals from 1000 hospital resamples of fixed predictions. Panel A shows the primary clinical comparisons; panel B shows the common-risk APACHE comparisons with training-fold recalibration. “Temporal addition” adds slope and rMSSD after mean and burden. Other rows compare the named additions with the panel baseline. All seven main-table comparisons are shown; full-refit ranges are in Table S11.

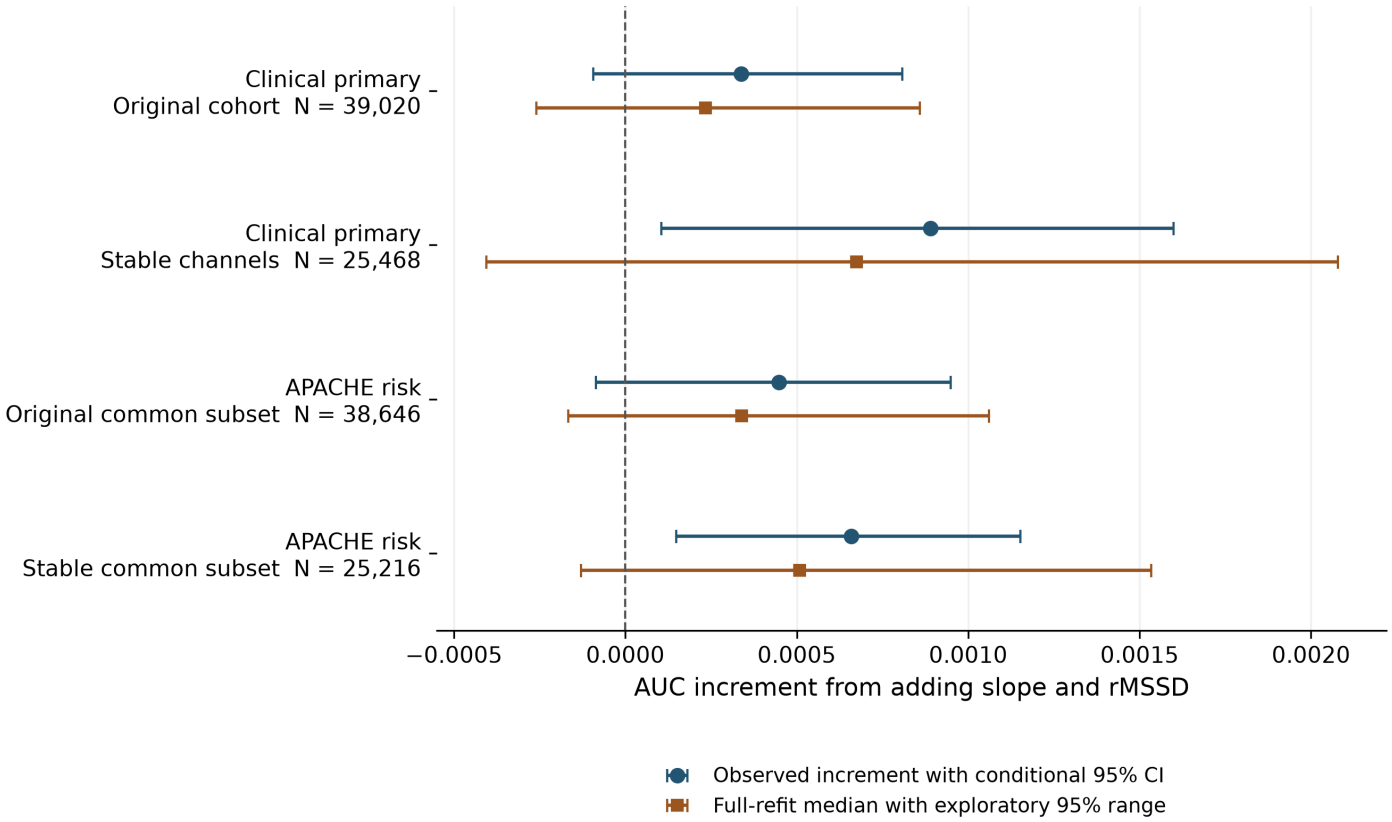


Figure 3

Temporal increments by channel availability

Observed AUC increments and conditional 95% intervals are shown separately from full-refit medians and exploratory 95% ranges. Temporal addition retains mean and burden and adds slope and rMSSD. The clinical and APACHE families are paired within each indicated population; contrasts between original and stable populations are descriptive. Positive conditional intervals in the stable subsets do not establish improvement in Brier score or clinical benefit.

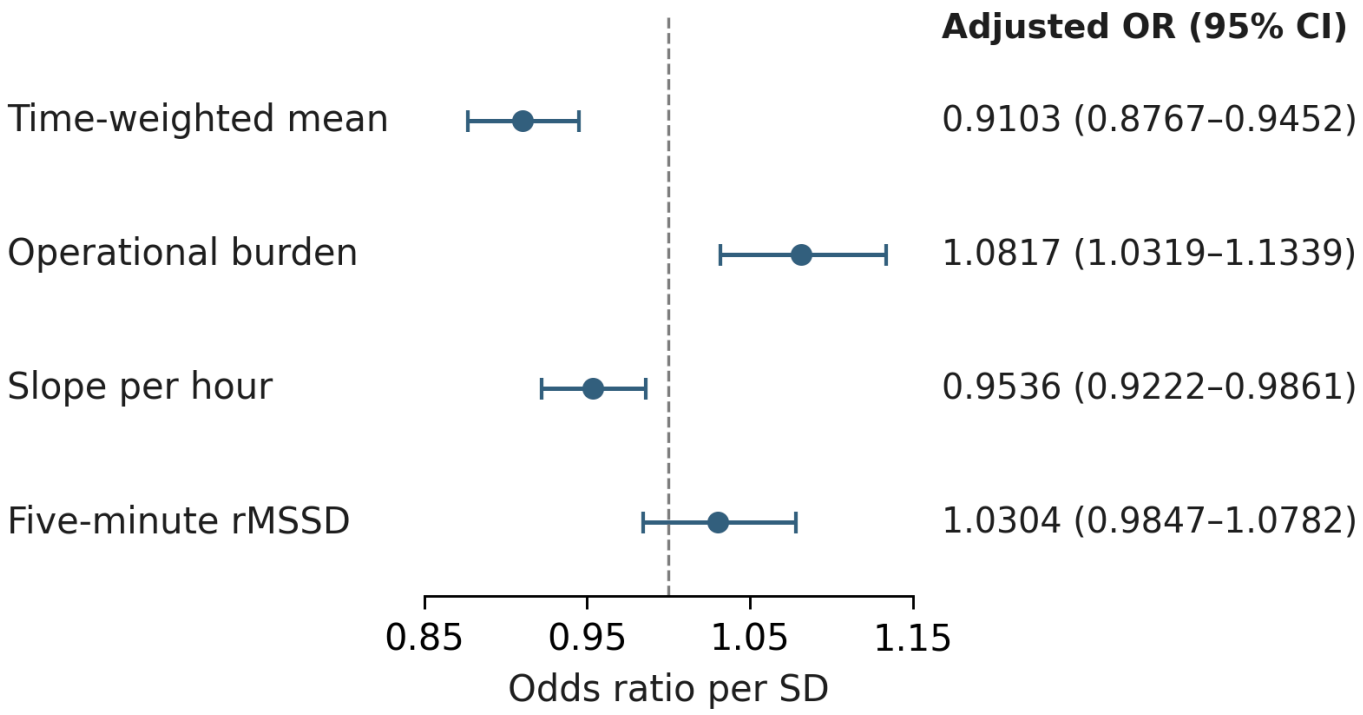


Figure 4

Joint adjusted associations

Odds ratios per cohort SD with hospital-clustered robust 95% CIs. All four ST summaries were included in the same clinically adjusted model. Mean, burden and slope showed associations with hospital mortality; the rMSSD interval included one. Full coefficients are in Table S6.

Supplementary Files

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

- Additionalfile3.zip
- Additionalfile2.zip
- Additionalfile1.docx