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.

Time in range and glucose trajectory phenotype, rather than glycemic variability alone, improve risk discrimination in sepsis: derivation in MIMIC-IV and external validation in the eICU Collaborative Research Database

Lei Wang

1453134107@qq.com

The First People's Hospital of Shunde

Lizhi Huang

The First People's Hospital of Shunde

Guojie Hu

The First People's Hospital of Shunde

Research Article

Keywords: Sepsis, Glycemic variability, Time in range, Group-based trajectory modeling, Stress hyperglycemia ratio, Prognosis, Incremental value, MIMIC-IV, External validation

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11070724/v1

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

Additional Declarations: No competing interests reported.

Time in range and glucose trajectory phenotype, rather than glycemic variability alone, improve risk discrimination in sepsis: derivation in MIMIC-IV and external validation in the eICU Collaborative Research Database

Authors' information

Author 1: Lei Wang, The Eighth Affiliated Hospital, Southern Medical University (The First People's Hospital of Shunde, Foshan). No. 1 Jiazi Road, Lunjiao, Shunde District, Foshan City, Guangdong Province, China.

Author 2: Lizhi Huang, The Eighth Affiliated Hospital, Southern Medical University (The First People's Hospital of Shunde, Foshan). No. 1 Jiazi Road, Lunjiao, Shunde District, Foshan City, Guangdong Province, China.

Author 3: Guojie Hu, The Eighth Affiliated Hospital, Southern Medical University (The First People's Hospital of Shunde, Foshan). No. 1 Jiazi Road, Lunjiao, Shunde District, Foshan City, Guangdong Province, China.

Authors' contributions

Author 1 and author 2 contributed equally to this work. Author 1 and author 3 are co-corresponding authors. Author 1 was responsible for study design, data extraction and manuscript writing. Author 2 performed statistical analysis, prepared figures and drafted the manuscript. Author 3 participated in study design, revised the manuscript and validated statistical methods. All authors reviewed the manuscript.

Corresponding author

Guojie Hu, 277426690@qq.com. The Eighth Affiliated Hospital, Southern Medical University (The First People's Hospital of Shunde, Foshan). No. 1 Jiazi Road, Lunjiao, Shunde District, Foshan City, Guangdong Province, China.

Lei Wang, 1453134107@qq.com. The Eighth Affiliated Hospital, Southern Medical University (The First People's Hospital of Shunde, Foshan). No. 1 Jiazi Road, Lunjiao, Shunde District, Foshan City, Guangdong Province, China.

Abstract

Background: Hyperglycemia and glycemic variability are common in sepsis and are linked to poor outcomes, but whether dynamic glucose measures carry prognostic information beyond established severity scores remains unclear. We tested whether a strong association with mortality also implies incremental discriminative value.

Methods: Retrospective cohort study in MIMIC-IV (version 3.1) with external validation in the eICU Collaborative Research Database (eICU-CRD). Adults meeting Sepsis-3 criteria were included. Variability was quantified by the coefficient of variation (CV), standard deviation, mean amplitude of glycemic excursion (MAGE), time in range (TIR; 70-180 mg/dL) and range; SHR was admission glucose divided by HbA1c-derived estimated average glucose. GBTM used cubic polynomials. The primary outcome was 28-day mortality. Analyses included multivariable regression, restricted cubic splines, time-varying Cox models, propensity matching and incremental discrimination testing.

Results: Among 6,290 MIMIC-IV patients, 28-day mortality was 23.8%. All variability measures were independently associated with death: CV (adjusted OR 1.022 per 1%), MAGE (OR 1.003 per 1 mg/dL) and TIR (OR 0.970 per 1%), all $P < 0.0001$. GBTM identified five trajectories spanning 13.0% to 51.3% mortality (very high-stable vs low-declining adjusted OR 4.28, 95% CI 2.81-6.54), and the very high-stable group retained a two-fold hazard after propensity matching (HR 2.21, 95% CI 1.52-3.19). Association and discrimination nevertheless diverged: adding CV to a base severity model raised the AUC by 0.005 (95% CI 0.001 to 0.008; $P = 0.004$) against 0.030 for TIR and 0.023 for the trajectory group (both $P < 0.0001$) and did not improve the integrated discrimination improvement (0.009; $P = 0.31$); combining CV with TIR and MAGE did not improve on TIR alone (AUC 0.8323 versus 0.8320). In 15,545 eICU-CRD patients the five-class structure was recovered (average posterior probability 0.882), the MIMIC-IV model transported with a mean curve difference of 6.6 mg/dL, and TIR again outperformed CV.

Conclusions: Glycemic variability is a strong marker of illness severity in sepsis but adds almost nothing to severity-based risk prediction, whereas time in range and glucose

trajectory phenotype carry independent prognostic information. Risk stratification should therefore favour time in range over variability indices.

Keywords: Sepsis; Glycemic variability; Time in range; Group-based trajectory modeling; Stress hyperglycemia ratio; Prognosis; Incremental value; MIMIC-IV; External validation

Background

Sepsis is a life-threatening condition caused by a harmful body response to infection [1]. It remains a leading cause of death in intensive care units (ICUs) worldwide [2, 3], and hospital mortality is still between 20% and 30% [3, 4]. The Sepsis-3 criteria, which combine suspected infection with an acute rise in the Sequential Organ Failure Assessment (SOFA) score, were validated against mortality in large cohorts [5, 6], and current guidelines recommend early risk stratification [7]. Better tools for predicting outcome and guiding treatment are still needed.

High blood glucose, or stress hyperglycemia, is common in sepsis. It follows the release of stress hormones and inflammatory mediators that raise blood glucose [8, 9]. Hyperglycemia in sepsis has been linked to higher death rates, longer stays and more complications [8, 10], and the association holds across admission diagnoses [11, 12]. The average glucose level is not the whole story. How much glucose swings during the ICU stay may matter just as much.

Blood glucose variability describes how much glucose moves up and down over time. It is usually measured by the coefficient of variation (CV), the standard deviation (SD), or the mean amplitude of glycemic excursion (MAGE) [13]. Large cohort studies have linked greater variability to higher mortality in critically ill patients [14-18], and the association persists after adjustment for mean glucose and illness severity [16, 18]. Time in range (TIR) is a related measure that captures how long glucose stays inside a target window and also tracks with survival [19]. Whether variability causes harm or simply marks severity is still debated [20, 21]. These studies share one limitation: each patient is reduced to a single summary number, and the day-to-day pattern of glucose is not examined.

The stress hyperglycemia ratio (SHR) combines the admission glucose with the patient's usual glucose level, estimated from HbA1c [22]. It separates true stress hyperglycemia from chronic hyperglycemia in diabetes. SHR predicts death better than admission glucose in several critically ill populations [23], and a dose-response meta-analysis confirmed the association across critically ill cohorts [24]. Sepsis-specific evidence is growing; a cohort of 13,199 patients with sepsis reported an independent association between SHR and mortality [24]. The shape of this association is still debated, with both J-shaped and U-shaped patterns reported [23, 24].

Group-based trajectory modeling (GBTM) sorts patients into groups that follow similar patterns over time [25, 26]. It has been used to describe kidney function after cardiac surgery [27] and growth patterns in children [28]. In glucose research, GBTM identifies patients with different glucose patterns during the ICU stay. This is more informative than a single average because it keeps the time dimension. Reporting standards for latent class trajectory models have been published [29].

Recent studies have applied GBTM or cluster analysis to blood glucose in sepsis [30-32]. These studies generally used three-group models or reported clusters on their own, without combining them with other glycemic measures or with methods that address confounding by illness severity. Two questions have therefore not been answered. First, does the day-to-day glucose pattern carry information that standard variability measures miss? Second, and more important for clinical use, do these measures improve risk prediction beyond the severity scores that are already available at the bedside, or do they simply mark how sick a patient is? We addressed both in one cohort: we (1) identified distinct blood glucose trajectory groups in patients with sepsis using GBTM with a five-group model; (2) tested the link between these groups and 28-day death; (3) examined traditional variability measures (CV, SD, MAGE, TIR) and SHR using restricted cubic splines and time-varying Cox regression; and (4) tested whether glycemic measures, individually and as trajectory groups, improve discrimination and reclassification beyond the SOFA score and beyond a full set of severity scores, with propensity score matching used to cross-validate the findings.

Methods

Data source

This study used the Medical Information Mart for Intensive Care IV (MIMIC-IV, version 3.1) database [33]. MIMIC-IV contains de-identified health records of patients admitted to the ICU at Beth Israel Deaconess Medical Center in Boston, USA, from 2008 to 2022. The database is publicly available to researchers who complete the required training and sign a data use agreement; the first author completed the Collaborative Institutional Training Initiative programme and is a certified PhysioNet user (record ID: 78245640). Data were extracted with Structured Query Language (SQL) using PostgreSQL (version 16). The Institutional Review Board of Beth Israel Deaconess Medical Center approved the MIMIC-IV database collection, and the requirement for informed consent was waived because the data was de-identified. The eICU Collaborative Research Database was collected under the oversight of the Institutional Review Board at the Beth Israel Deaconess Medical Center[34]; all data are de-identified and its use for research is covered by the PhysioNet credentialed-access agreement.

Study population

We included adult patients (aged 18 years or older) who met the Sepsis-3 criteria [1]. Sepsis-3 was defined as suspected infection combined with an acute increase in SOFA score of 2 points or more [1, 5]. The SOFA score was calculated as originally described [35]. We excluded patients with ICU stays shorter than 24 hours and those with fewer than 3 blood glucose measurements during their ICU stay. When a patient had multiple ICU stays, only the first stay was included.

Blood glucose data extraction

We extracted blood glucose readings from laboratory tests (itemids 50809 and 50931; blood-gas and chemistry glucose) and from ICU bedside charting records (itemids 220621, 225664 and 226537). All values were in mg/dL. We combined these data sources and removed readings outside the range of 10 to 2000 mg/dL (likely errors). For each patient, we calculated the following measures during the ICU stay:

- **Mean glucose:** average of all blood glucose readings

- **CV (%):** (SD / mean) x 100
- **SD:** standard deviation of all blood glucose readings
- **MAGE:** mean of blood glucose swings greater than 1 SD [13]
- **TIR (%):** percentage of readings within 70 to 180 mg/dL
- **Range:** maximum minus minimum blood glucose
- **Hypo episodes:** number of readings below 70 mg/dL
- **Hyper episodes:** number of readings above 180 mg/dL

SHR calculation

SHR was defined as the first ICU blood glucose reading divided by the estimated average glucose (eAG). The eAG was calculated from HbA1c using the ADAG study formula: eAG (mg/dL) = 28.7 x HbA1c (%) - 46.7 [36]. HbA1c was taken from the closest lab measurement within 48 hours of ICU admission. Patients without HbA1c data were excluded from SHR analysis but remained in the main analysis.

Outcomes

The primary outcome was 28-day death, defined as death within 28 days of ICU admission. We also recorded 90-day and 365-day death as secondary outcomes. Survival time was calculated from ICU admission to death or last follow-up, truncated at 28 days for the main analysis.

Group-based trajectory modeling (GBTM)

We used latent class growth analysis (LCGA), a special case of group-based trajectory modeling in which the within-class random effects are fixed at zero, to identify groups of patients following similar daily blood glucose trajectories [25]. For each patient we computed the mean of all glucose measurements recorded on each ICU day, from the day of admission (day 0) through day 14, and included patients with glucose recorded on at least 3 separate days. Only measurements taken during the index hospital admission were used: values were retained when their admission identifier matched the index admission, or, where that identifier was missing, when the sampling time fell between hospital admission and discharge. This excluded 1,716 of 343,462 values inside the day 0-14 window (0.50%, from 390 patients) that were recorded during a different admission, and

left 5,757 of the 6,290 cohort patients with at least 3 eligible days. Days with no measurement were treated as missing and contributed no information, so the likelihood was evaluated only over observed days and patients with incomplete series were retained.

The trajectory of each class was modeled as a polynomial function of ICU day, and a single residual variance was estimated and shared across all classes. This is the specification used by the two established implementations of the method, the Stata *traj* plugin and the R package *lcmm* when the residual variance is constrained to be common across classes [37, 38]. The trajectory solution reported here is nevertheless our own: it was fitted with a purpose-written maximum-likelihood routine in Python, and it was subsequently checked against the R implementation rather than assumed to be correct (see Statistical analysis). This choice of variance specification matters for model comparison: because the Bayesian Information Criterion (BIC) is computed from the likelihood, a specification that instead estimates a separate residual variance for each class inflates the likelihood and yields systematically lower BIC values. All BICs reported here are mutually comparable, but not comparable with values obtained under a class-specific-variance specification. We fitted a complete grid of 18 models comprising 2 to 7 classes crossed with polynomial degrees of 1 (linear), 2 (quadratic) and 3 (cubic). Parameters were estimated by maximum likelihood using an expectation-maximization (EM) algorithm run from 20 random starts plus one deterministic start that ranked patients by mean glucose; the solution with the highest log-likelihood was retained.

Model selection followed pre-specified criteria. BIC = $-2 \times$ log-likelihood + $p \times \ln N$, where p is the number of estimated parameters and N the number of patients. Because BIC continues to fall as classes are added in large samples, a larger model was preferred only if it also met all of the following: average posterior probability of assignment (AvePP) of at least 0.70 for every class [25, 26]; odds of correct classification (OCC) above 5 for every class; a smallest class containing at least 1% of the sample; and clinical distinctiveness, meaning that no new class differed from its nearest neighbouring class by less than 5 percentage points in 28-day mortality. Each patient was assigned to the class with the highest posterior probability.

To confirm that the selected solution did not depend on our own implementation, we refitted all 18 models with an implementation written independently in a different software environment: R (version 4.6.1) and its *lcmm* package (version 2.2.2; function *hlme* fitted without random effects), which estimates the identical latent class growth model with a different optimiser. At five classes the two independent implementations gave log-likelihoods of −245,472.3 (Python) and −245,472.6 (R), agreeing to within 0.0002% and both far better than every smaller model. The small residual difference reflects the well-known multi-modality of the mixture likelihood: from its own default starting values the R fit reached a marginally lower optimum (−245,472.6) whose partition agreed with the retained one for 87% of patients (label-matched Cohen's κ = 0.75). To distinguish this from any genuine disagreement between the two programs, we passed the retained Python estimates to the R fit as starting values. *lcmm* evaluated the log-likelihood at those values as −245,472.3009, identical to the Python value to four decimal places, and its optimiser did not leave the point (largest change in any of the 25 parameters 0.0009, converged log-likelihood −245,472.298). It reproduced the identical BIC (491,161.1) and assigned 99.95% of patients to the same class (κ = 0.999), with class sizes differing by at most two patients and the mortality gradient unchanged (13.0%, 23.6%, 32.5%, 51.0% and 33.5% across groups 1 to 5). The two implementations therefore maximise the same likelihood, and the retained solution is a stable optimum of it. Restarting the EM algorithm from 400 random initialisations across five seeds returned this same optimum every time.

Statistical analysis

Descriptive statistics and missing data. Continuous variables are reported as median with interquartile range (IQR) and compared using the Mann-Whitney U test for two-group comparisons and the Kruskal-Wallis test for comparisons across the five trajectory groups; categorical variables are reported as number (%) and compared using the chi-square test or Fisher's exact test, as appropriate. The glycemic exposure variables (mean glucose, CV, SD, MAGE, TIR) were complete for all 6,290 patients, as were all covariates. HbA1c, and with it SHR, was available for 1,646 patients (26.2%); these patients were analysed as a defined subgroup, with complete-case analysis as the primary

approach and multiple imputation as a sensitivity analysis. No a priori sample size calculation was performed; the cohort comprised all eligible admissions in the database.

Outcome and follow-up. The primary outcome was death within 28 days of ICU admission. Follow-up started at ICU admission and ended at death or at day 28, whichever came first; patients still alive were censored at day 28. Death at 90 days and at 365 days were pre-specified secondary outcomes. Vital status was ascertained from hospital records and linked death registry data, and no patient was lost to follow-up for the 28-day outcome; ascertainment of the 90- and 365-day outcomes beyond hospital discharge depends on those linked records, which may miss deaths occurring outside the reporting system.

Multivariable models. We fitted a separate multivariable logistic regression model for each glycemic measure rather than entering them simultaneously, because the measures are strongly collinear (variance inflation factors 10.2 to 22.5). Every model was adjusted a priori for age, sex, diabetes status and SOFA score. Results are reported as adjusted odds ratios (ORs) with 95% confidence intervals (CIs). We fitted corresponding Cox proportional hazards models with the same covariates, reported as hazard ratios (HRs) with 95% CIs.

Non-linearity. We used restricted cubic splines with four knots placed at the 5th, 35th, 65th and 95th percentiles of each exposure to examine the shape of the exposure-response relationship [39]. Departure from linearity was tested by comparing the spline model with the corresponding linear model using a likelihood ratio test; we report both the overall and the non-linear *P* value.

Time-varying Cox regression. To use the repeated daily glucose measurements we built a counting-process dataset in the (start, stop] format, in which each patient contributed one record per ICU day and daily mean glucose and daily CV entered as time-varying covariates. Because exposure is updated as follow-up proceeds, this specification avoids immortal-time bias and estimates the separate contributions of the glucose level and of glucose variability measured on the same day.

Propensity score matching. Matching was performed in the full cohort of 6,290 patients. Patients were classified as high-variability (CV above the median of 23.6%) or low-

variability (CV at or below the median). Propensity scores were estimated by logistic regression including age, sex, diabetes, SOFA score, APS III [40], SAPS II and the Charlson comorbidity index [41]. We performed 1:1 nearest-neighbour matching on the logit of the propensity score without replacement [42], using a caliper of 0.1 standard deviations of the logit propensity score; treated patients with no eligible control inside the caliper were not matched. A caliper was mandatory here because, without one, every treated patient is matched and covariate balance cannot improve. Balance was assessed with standardized mean differences (SMD), with an absolute SMD below 0.1 indicating adequate balance; significance tests after matching are not reported because they are not informative for balance. Outcomes in the matched sample were compared with the log-rank test and with Cox regression using a robust (pair-clustered) variance estimator. Two pre-specified sensitivity analyses varied the algorithm: a wider caliper of 0.2 SD, and matching with replacement. In a second, pre-specified matching analysis, patients in the very high-stable trajectory group were matched to patients in all other groups; because diabetes was present in 83.3% of that group versus 30.5% of the remaining classes, matching was performed exactly on diabetes status and then 1:1 on the propensity score estimated from age, sex, diabetes and SOFA score.

Effect modification by diabetes. We repeated the principal analyses in patients with and without diabetes and tested for interaction by adding a product term between each glycemic measure and diabetes status.

Model diagnostics. The proportional hazards assumption was assessed with Schoenfeld residuals plotted against rank-transformed time, tested globally and for each covariate. Because the assumption was violated for age in every model, we addressed it in three pre-specified ways: (i) time-varying Cox models, in which the (start, stop] formulation does not require the exposure effect to be constant; (ii) Cox models stratified on age category (<65, 65–79 and ≥80 years), which removes the proportional hazards assumption for age entirely; and (iii) a landmark analysis. The stratified and unstratified estimates are reported side by side.

Incremental prognostic value. To test whether glycemic information adds to routine risk assessment rather than merely restating illness severity, we fitted a series of nested

multivariable logistic regression models, estimated by unpenalised maximum likelihood, for 28-day mortality and compared their discrimination in the full cohort of 6,290 patients. The reference model contained age, sex, diabetes status and the SOFA score. The glycemic measures (CV, TIR and MAGE) were then added, separately and jointly. A second reference model additionally contained APS III, SAPS II and the Charlson comorbidity index, representing a fuller severity adjustment, and the glycemic measures were added to it in the same way. Discrimination was summarised by the area under the receiver operating characteristic curve (AUC) with 95% confidence intervals from 2,000 bootstrap resamples; optimism was estimated by 200 bootstrap replications of the whole modelling procedure and subtracted from the apparent AUC to give an optimism-corrected value. Nested models were compared with the DeLong test for correlated curves and, where stated, with the likelihood ratio test. Because a change in the AUC can be statistically significant at a magnitude that is irrelevant clinically, we treated the integrated discrimination improvement, which additionally requires the average predicted risk to move in the right direction, as a co-primary measure of added predictive ability. The added predictive ability was further quantified with the continuous and the category-based net reclassification improvement (NRI) and with the integrated discrimination improvement (IDI), using the observed 28-day mortality (23.8%) as the single risk threshold for the categorical NRI [43], and clinical usefulness was compared by decision curve analysis across threshold probabilities from 5% to 60% [44]. Calibration was summarised by the Brier score and by the calibration slope and intercept obtained from the same 200 bootstrap replications used for the optimism correction. In a secondary analysis restricted to the 5,757 patients in the trajectory cohort, the five trajectory groups were added to the reference model to test whether the trajectory phenotype carried information beyond the individual measures.

Sensitivity analyses. We (i) expanded the covariate set to include APS III and the Charlson comorbidity index; (ii) adjusted all variability models for the total number of glucose measurements, because patients monitored more frequently have greater opportunity to exhibit variability (Spearman $\rho = 0.28$ between measurement count and CV); (iii) performed a landmark analysis in the full cohort restricted to patients still alive at day 7, to separate early from late effects and to address non-proportional hazards; (iv)

applied a Bonferroni correction to the six primary glycemic measures (adjusted $\alpha = 0.0083$); (v) repeated the SHR analysis in 10 multiply imputed data sets, pooling estimates with Rubin's rules; (vi) refitted the trajectory model with quadratic rather than cubic polynomials; (vii) quantified survivor bias in the trajectory cohort, which required at least 3 days of glucose data; and (viii) repeated all principal Cox models stratified on age category and, in a further step, simultaneously on age and SOFA category, to address violation of the proportional hazards assumption. Each of these analyses returns a single effect estimate for a defined model, and they are therefore reported together rather than as separate figures.

Marginal structural model. Insulin titration is a response to preceding glucose values and is therefore a time-varying confounder that is itself affected by earlier glucose, a structure that ordinary regression cannot handle. We therefore fitted a marginal structural model with stabilised inverse-probability-of-treatment weights. Exposure and the time-varying confounders were measured on each ICU day from day 0 to day 13, and death was modelled as an event occurring during the following day, so that exposure always preceded the outcome. The numerator of the treatment weight was conditional on baseline age, sex, diabetes and SOFA score, the previous day's exposure and study day, and the denominator additionally on the previous day's mean glucose, coefficient of variation and number of glucose measurements; a parallel weight accounted for remaining under ICU observation within the window. Weights were stabilised, truncated at the 1st and 99th percentiles, and the outcome was fitted by weighted pooled logistic regression, with the study day entered as a set of indicator variables and standard errors clustered by patient. Two exposures were modelled: insulin on the day, and blood glucose outside 70–180 mg/dL on the day. The results are given in Supplementary file 1, Table S7.

All tests were two-sided, with a significance level of $P < 0.05$.

Insulin exposure

Insulin is the dominant determinant of both glucose level and glucose variability in the ICU, and it is simultaneously a response to hyperglycemia and a cause of hypoglycemia, which makes it a time-varying confounder influenced by previous glucose values. We

therefore ascertained insulin exposure in the development cohort from two sources in MIMIC-IV v3.1. The first was *icu/inputevents*, using all items whose label contains insulin (itemid 223257 Insulin 70/30, 223258 Insulin Regular, 223259 Insulin NPH, 223260 Insulin Glargine, 223261 Insulin Humalog 75/25, 223262 Insulin Humalog, 229299 Insulin Novolog and 229619 Insulin U500; all amounts are recorded in units). The second was *hosp/prescriptions*, using every order whose drug name contains insulin, which captures subcutaneous sliding-scale regimens that are not always charted in the ICU flowsheets. Both sources were restricted to the index hospitalisation by matching *subject_id* and *hadm_id*. Day 0 was the ICU admission time and exposure was retained for days 0 to 14, the same window used for the trajectory model and the glucose metrics. The primary exposure was any insulin on any of days 0 to 14. We additionally recorded whether insulin was administered in the ICU, whether it was given as a continuous infusion (administration lasting 5 minutes or longer) or as intermittent injection, the number of days with exposure, the total number of units and the mean daily dose among exposed patients, and exposure restricted to the first seven days.

Insulin exposure accrues over time, so a patient who dies early has fewer opportunities to receive it, and a naive adjustment would bias the estimate in favour of treatment. We therefore addressed confounding by insulin in four ways. First, we added any exposure to the multivariable models. Second, we used a seven-day landmark analysis restricted to the patients still alive and under observation on day 7, with exposure limited to days 0 to 7 so that it precedes the landmark. Third, we fitted a time-varying Cox model in which insulin exposure was updated on each study day, which respects the temporal ordering of exposure and outcome. Fourth, we stratified all analyses by exposure and tested the interaction formally. All four approaches are reported in Supplementary file 1, together with the corresponding estimates for eICU-CRD.

External validation in the eICU Collaborative Research Database

To test whether the trajectory structure and the prognostic associations were reproducible and transportable outside the development database, all analyses were repeated in the eICU Collaborative Research Database (eICU-CRD, version 2.0) [34], a multicentre

registry of 200,859 ICU admissions from 208 hospitals in the United States. Eligibility criteria, glucose cleaning rules and variability formulas were identical to those used for MIMIC-IV. Three features of the source data are inherent to the database and are stated explicitly. First, eICU-CRD stores only offsets in minutes from ICU admission rather than absolute timestamps, so the analysis day was defined as floor(offset / 1,440). Second, the database contains no HbA1c measurement; the stress hyperglycemia ratio could therefore not be validated and is not reported for this cohort. Third, no 28-day vital status is available, so the outcomes were hospital and ICU mortality. Glucose values came from the clinical laboratory (labname 'glucose') and from bedside point-of-care testing (nurseCharting 'Bedside Glucose'). Suspected infection was identified from Angus ICD-9 infection codes or from the structured diagnosis string, and baseline SOFA was computed from the day-1 APACHE IV variables (PaO2/FiO2 ratio, bilirubin, mean arterial pressure, vasopressor use, Glasgow Coma Scale, creatinine and urine output) together with the platelet nadir.

Two complementary external analyses were performed. First, the model was refitted de novo: the same grid (2 to 7 classes, polynomial degrees 1 to 3, single common residual variance) was re-estimated in eICU-CRD and the recovered mean curves were compared with the MIMIC-IV curves after optimal one-to-one matching on the root-mean-square difference. Second, transportability was assessed: the MIMIC-IV parameter estimates (class coefficients, mixing proportions and residual variance) were frozen and applied to eICU-CRD patients, who were then assigned to the class with the highest posterior probability under this imported model; the average posterior probability of assignment, the resulting class proportions and the mortality gradient were examined. In addition, all multivariable models (logistic regression for hospital and ICU mortality, discrimination and decision-analytic measures) were re-estimated in eICU-CRD, and insulin exposure was ascertained in *both* cohorts and introduced as a covariate in parallel sensitivity analyses (see Insulin exposure).

Software and reproducibility. The trajectory model was fitted in Python (version 3.13) with a purpose-written expectation-maximisation routine, and all other estimates (regression, spline, propensity-matching, discrimination and decision-analytic statistics) were computed in Python with the pandas, statsmodels, lifelines and scikit-learn

packages; the same code was applied to the external cohort with eICU-CRD inputs. Because the trajectory solution is therefore a product of our own implementation, it was verified against an established implementation rather than assumed to be correct: the same 18 models were refitted in R (version 4.6.1) with the lcmm package, and every other reported estimate was reproduced independently in R with the survival and splines packages. The log-likelihoods quoted for the trajectory models are the Python values, which the R fit reproduced to four decimal places; the two environments agreed to the reported precision for all remaining estimates.

Use of large language models. Large language models were used only to assist with language editing and with drafting and debugging of analysis code. They did not contribute to the study design, the data analysis or the interpretation of the results, and all authors reviewed and verified the content of the manuscript and take full responsibility for it. The study is reported in accordance with the STROBE statement [45].

Results

Study population

A total of 6,290 sepsis patients were included. The median age was 68.4 years, 61.2% were male, and 31.6% had diabetes. The median SOFA score was 9, and the median ICU stay was 3.4 days. The 28-day death rate was 23.8% (1,494 of 6,290 patients). Patients who died were older (70.3 vs 68.1 years, $P < 0.0001$), had higher SOFA scores (12.5 vs 8.0, $P < 0.0001$), and had longer ICU stays (4.1 vs 3.3 days, $P < 0.0001$) (Table 1, Figure 1).

Table 1. Baseline characteristics of the study population by 28-day mortality

Variable	Dead (n=1,494)	Alive (n=4,796)	P value
Age, median (IQR)	70.3 (57.7-81.2)	68.1 (58.4-77.0)	<0.0001
Male, n (%)	853 (57.0%)	2,997 (62.5%)	0.0001
Diabetes, n (%)	464 (31.1%)	1,523 (31.7%)	0.68
SOFA score, median (IQR)	12.5 (9.0-16.0)	8.0 (5.0-10.0)	<0.0001
APSIII, median (IQR)	76.0 (58.0-97.0)	44.0 (32.0-65.0)	<0.0001
SAPSII, median (IQR)	56.0 (46.0-67.0)	40.0 (34.0-51.0)	<0.0001

Charlson score, median (IQR)	6.0 (4.0-8.0)	4.0 (3.0-6.0)	<0.0001
ICU stay (days), median (IQR)	4.1 (2.1-8.1)	3.3 (1.8-6.9)	<0.0001
First glucose (mg/dL), median (IQR)	153 (112-222)	143 (117-181)	<0.0001
Mean glucose (mg/dL), median (IQR)	145 (121-181)	131 (121-146)	<0.0001
CV (%), median (IQR)	28.6 (20.6-38.4)	22.6 (17.8-29.3)	<0.0001
MAGE (mg/dL), median (IQR)	65.1 (43.5-97.0)	46.8 (35.2-66.3)	<0.0001
TIR (%), median (IQR)	76.0 (50.0-91.7)	92.3 (80.0-100.0)	<0.0001
SHR, median (IQR)	1.32 (0.98-1.80)	1.12 (0.92-1.37)	<0.0001
HbA1c (%), median (IQR)	6.0 (5.4-6.9)	5.9 (5.6-6.7)	0.44
Any glucose < 70 mg/dL, n (%)	550 (36.8%)	1,035 (21.6%)	<0.0001
Any glucose < 54 mg/dL, n (%)	130 (8.7%)	91 (1.9%)	<0.0001
Any insulin exposure, days 0-14, n (%)	1,103 (73.8%)	4,199 (87.6%)	<0.0001
Insulin administered in the ICU, n (%)	961 (64.3%)	3,955 (82.5%)	<0.0001
Days with insulin exposure, median (IQR)	3.0 (0.0-7.0)	6.0 (5.0-10.0)	<0.0001

Blood glucose variability and 28-day death

All blood glucose variability measures were strongly linked to 28-day death in both unadjusted and adjusted models (Table 2, Figure 2). After adjusting for age, sex, diabetes, and SOFA score, each 1% increase in CV was linked to a 2.2% higher odds of death (adjusted OR 1.022, 95% CI 1.016 to 1.028, $P < 0.0001$). Each 1% increase in TIR was linked to a 3.0% lower odds of death (adjusted OR 0.970, 95% CI 0.966 to 0.973, $P < 0.0001$). MAGE was also linked to death (adjusted OR 1.003 per 1 mg/dL, $P < 0.0001$).

Hypoglycemia was common and tracked with outcome. At least one glucose reading below 70 mg/dL occurred in 1,585 patients (25.2%) and at least one below 54 mg/dL in 221 patients (3.5%). Both were more frequent in patients who died (below 70 mg/dL, 550

of 1,494 [36.8%] versus 1,035 of 4,796 [21.6%]; below 54 mg/dL, 130 [8.7%] versus 91 [1.9%]; both $P < 0.0001$; Table 1).

RCS analysis showed non-linear relationships for CV, mean glucose, MAGE, and TIR (all non-linear $P < 0.0001$; Figure 3). The risk of death rose more steeply at higher levels of these measures. The curves revealed U-shaped relationships for CV and mean glucose, with the lowest mortality risk at CV around 22% and mean glucose around 125 mg/dL, while TIR showed an inverse pattern with risk falling as TIR increased up to approximately 85%. Correlations among the variability measures are shown in Figure 4.

Table 2. Multivariable logistic regression for blood glucose variability measures and 28-day death

Measure	Adjusted OR	95% CI	P value	P (Bonferroni)*
Mean glucose (per 1 mg/dL)	1.013	1.012-1.015	<0.0001	<0.0006
CV (per 1% increase)	1.022	1.016-1.028	<0.0001	<0.0006
SD (per 1 mg/dL)	1.017	1.014-1.020	<0.0001	<0.0006
MAGE (per 1 mg/dL)	1.003	1.002-1.004	<0.0001	<0.0006
TIR (per 1% increase)	0.970	0.966-0.973	<0.0001	<0.0006
Range (per 1 mg/dL)	1.000	1.000-1.001	0.16	0.96 (NS)

*Adjusted for age, sex, diabetes, and SOFA score. Bonferroni correction applied for 6 primary comparisons (α=0.0083). OR = odds ratio; CI = confidence interval; CV = coefficient of variation; SD = standard deviation; MAGE = mean amplitude of glycemic excursion; TIR = time in range (70-180 mg/dL); NS = not significant.

SHR and death

Among the 1,646 patients with HbA1c data, SHR was higher in those who died (median 1.32 vs 1.12, $P < 0.0001$). After adjusting for age, sex, diabetes, and SOFA score, SHR was an independent predictor of 28-day death (adjusted OR 1.026 per 0.1 increase, 95% CI 1.007 to 1.048, $P = 0.013$; Table 3). Cox regression confirmed this finding (adjusted HR 1.017 per 0.1 increase, 95% CI 1.006 to 1.027, P = 0.003; Table 3). RCS analysis of

SHR revealed a non-linear, J-shaped relationship with mortality (non-linear $P = 0.014$; Figure 5), with risk rising steeply when SHR exceeded approximately 1.5.

When patients were divided into SHR tertiles, those in the highest tertile (SHR > 1.3) had a 19.5% death rate compared with 10.2% in the lowest tertile (SHR < 1.0). The adjusted OR for the highest versus lowest tertile was 1.69 (95% CI 1.16 to 2.48, $P = 0.007$, P for trend <0.001).

Table 3. SHR tertile analysis and 28-day death

SHR Tertile	**N**	**Deaths (%)**	**Mean SHR**	**Adjusted estimate (95% CI)**	**P value**
T1 (Low, <1.0)	549	56 (10.2%)	0.81	Reference	-
T2 (Mid, 1.0-1.3)	548	45 (8.2%)	1.14	0.87 (0.56-1.35)	0.53
T3 (High, >1.3)	549	107 (19.5%)	1.84	1.69 (1.16-2.48)	0.007
Continuous SHR, logistic regression					
SHR per 0.1 increase, complete case	1,646	208 (12.6%)	1.14	1.026 (1.007-1.048)	0.013
SHR per 0.1 increase, multiple imputation (m = 10)	6,290	1,494 (23.8%)	-	1.040 (1.022-1.058)	<0.0001
Continuous SHR, Cox regression					
SHR per 0.1 increase, complete case	1,646	208 (12.6%)	1.14	1.017 (1.006-1.027)	0.003
SHR per 0.1 increase, multiple imputation (m = 10)	6,290	1,494 (23.8%)	-	1.024 (1.018-1.030)	<0.0001

Adjusted for age, sex, diabetes, and SOFA score. P for trend <0.001. Estimates in the logistic rows are odds ratios and estimates in the Cox rows are hazard ratios, all

expressed per 0.1 increase in SHR and all adjusted for the same covariates; the Cox models use the same 28-day time scale. The complete-case rows use the 1,646 patients with a measured HbA1c and the imputed rows pool m = 10 completed data sets with Rubin's rules (fraction of missing information 0.63). SHR = stress hyperglycemia ratio.

Blood glucose trajectory patterns (GBTM)

We included 5,757 patients with at least 3 days of blood glucose data recorded during the index hospital admission. All 5,757 received a class assignment. BIC decreased monotonically as classes were added, reaching its lowest value at seven classes (Supplementary file 2, Table S8). The six- and seven-class models split existing classes into subgroups with nearly identical 28-day mortality (differences of less than 3 percentage points between the added class and its nearest neighbour) and had lower average posterior probabilities (0.86 and 0.84, versus 0.87 for five classes). Applying our pre-specified adequacy criteria, we selected the five-class cubic model (K5_D3: log-likelihood −245,472.3; BIC 491,161.1; AvePP 0.869; smallest class 156 patients) (Supplementary file 2, Figure 6). Every class in this model had a class-specific AvePP of at least 0.78 and an odds of correct classification far above 5.

BIC = Bayesian Information Criterion; AvePP = average posterior probability of assignment. All 18 fitted models are shown, ordered by BIC.

The five groups showed distinct blood glucose patterns (Table 4, Figure 7):

- **Group 1 (Low-Declining):** 3,739 patients (64.9%). Blood glucose started at 133 mg/dL and fell slowly to about 110 mg/dL. Mean glucose 126 mg/dL. Death rate 13.0%.
- **Group 2 (Moderate-Stable):** 1,132 patients (19.7%). Blood glucose stayed between about 140 and 156 mg/dL. Mean glucose 148 mg/dL. Death rate 23.7%.
- **Group 3 (High-Stable):** 460 patients (8.0%). Blood glucose stayed between about 165 and 200 mg/dL. Mean glucose 174 mg/dL. Death rate 32.4%.
- **Group 4 (Very High-Stable):** 156 patients (2.7%). Blood glucose stayed above 205 mg/dL throughout the ICU stay. Mean glucose 229 mg/dL. Death rate 51.3%.

- **Group 5 (High-Start Declining):** 270 patients (4.7%). Blood glucose started at 264 mg/dL and fell to 139 mg/dL by day 14. Mean glucose 208 mg/dL. Death rate 33.3%.

Table 4. Characteristics of the five blood glucose trajectory groups

Variable	**G1: Low-Declining**	**G2: Moderate-Stable**	**G3: High-Stable**	**G4: Very High-Stable**	**G5: High-Start Declining**	**P**
N (%)	3,739 (64.9%)	1,132 (19.7%)	460 (8.0%)	156 (2.7%)	270 (4.7%)	-
Age, median	68.3	69.6	68.4	68.5	67.1	0.66
Male (%)	61.1%	61.8%	62.6%	61.5%	60.4%	0.96
Diabetes (%)	17.2%	47.9%	71.1%	83.3%	72.2%	<0.0001
SOFA, median	8.0	9.0	10.0	11.0	11.0	<0.0001
HbA1c (%), median	5.8	6.5	7.3	8.0	7.3	<0.0001
SHR, median	1.1	1.1	1.2	1.5	1.8	<0.0001
Mean glucose (mg/dL)	126	148	174	229	208	<0.0001
CV (%), median	21.6	25.7	29.5	33.6	38.4	<0.0001
TIR (%), median	95.2	82.4	58.9	28.6	47.5	<0.0001
28-day death (%)	13.0%	23.7%	32.4%	51.3%	33.3%	<0.0001
90-day death (%)	16.6%	29.9%	40.7%	57.7%	41.5%	<0.0001
365-day death (%)	21.2%	36.0%	47.8%	62.2%	50.0%	<0.0001
Any glucose < 70 mg/dL, %	23.3%	27.7%	25.9%	27.6%	31.5%	0.002
Any insulin exposure, %	80.6%	91.5%	97.2%	100%	99.3%	<0.0001
Insulin dose, units/day among exposed, median	5.9	8.0	14.1	26.7	17.7	<0.0001

Mean glucose for each trajectory group is calculated at the patient level (the mean of each patient's ICU mean glucose), ensuring consistency between text and table values.

Kaplan-Meier curves differed across the five classes (overall log-rank $P = 1.7 \times 10^{-68}$; Figure 8). In pairwise comparisons with Bonferroni correction (threshold $P < 0.005$), nine of the ten class pairs differed significantly; the exception was group 3 versus group 5 (P = 0.57), which had almost identical 28-day mortality (32.4% versus 33.3%) despite clearly different trajectory shapes. The proportion of patients with at least one glucose reading below 70 mg/dL rose modestly across the groups (23.3%, 27.7%, 25.9%, 27.6% and 31.5% for Groups 1 to 5; $P = 0.002$; Table 4).

After adjusting for age, sex, diabetes, and SOFA score, all higher trajectory groups had significantly higher odds of death compared with Group 1 (Table 5). Group 4 (very high-stable) had the highest risk (adjusted OR 4.28, 95% CI 2.81 to 6.54, $P < 0.001$); the full set of odds ratios and hazard ratios is shown in Table 5 and Figure 9. Cox regression confirmed these findings (Table 5). The gradient persisted with longer follow-up: 90-day mortality was 16.6%, 29.9%, 40.7%, 57.7% and 41.5% across Groups 1 to 5, and 365-day mortality was 21.2%, 36.0%, 47.8%, 62.2% and 50.0% (both $P < 0.0001$; Table 4).

Table 5. Association between blood glucose trajectory groups and 28-day death

Comparison (vs G1)	**Unadjusted OR (95% CI)**	**P**	**Adjusted OR (95% CI)***	**P**	**Cox HR (95% CI)***	**P**
G2: Moderate-Stable	2.08 (1.76-2.45)	<0.001	1.81 (1.49-2.20)	<0.001	1.51 (1.30-1.77)	<0.001
G3: High-Stable	3.21 (2.58-3.99)	<0.001	3.21 (2.39-4.30)	<0.001	2.32 (1.90-2.83)	<0.001
G4: Very High-Stable	7.05 (5.07-9.78)	<0.001	4.28 (2.81-6.54)	<0.001	3.48 (2.69-4.50)	<0.001
G5: High-Start Declining	3.35 (2.55-4.39)	<0.001	1.91 (1.35-2.69)	<0.001	2.18 (1.71-2.77)	<0.001

*Adjusted for age, sex, diabetes, and SOFA score. Odds ratios are from pairwise models contrasting each class with G1; hazard ratios are from a single Cox model containing all

five classes. Unadjusted odds ratios are identical under the two parameterisations. OR = odds ratio; HR = hazard ratio; CI = confidence interval.

Time-varying Cox regression

To account for the fact that blood glucose changes over time, we built time-varying Cox models using daily blood glucose measures (Table 6). In the joint model including both daily mean glucose and daily CV, both were independent predictors of death: daily mean glucose (HR 1.005 per 1 mg/dL, 95% CI 1.003 to 1.006, $P < 0.0001$) and daily CV (HR 1.015 per 1%, 95% CI 1.010 to 1.020, $P < 0.0001$). This means that both the blood glucose level and how much it swings on a given day independently affect the risk of death.

Table 6. Time-varying Cox regression for daily blood glucose measures and death

Variable	HR	95% CI	P value
Model 1: Daily mean glucose only			
Daily mean glucose (per 1 mg/dL)	1.005	1.004-1.006	<0.0001
Model 2: Daily CV only			
Daily CV (per 1%)	1.018	1.013-1.023	<0.0001
Model 3: Joint model (both)			
Daily mean glucose (per 1 mg/dL)	1.005	1.003-1.006	<0.0001
Daily CV (per 1%)	1.015	1.010-1.020	<0.0001
Model 4: Joint model with daily insulin			
Daily mean glucose (per 1 mg/dL)	1.005	1.004-1.006	<0.0001
Daily CV (per 1%)	1.016	1.011-1.021	<0.0001
Insulin exposure on the same day	0.735	0.622-0.867	0.0003

All models adjusted for age, sex, diabetes, and SOFA score. HR = hazard ratio; CI = confidence interval; CV = coefficient of variation.

Propensity score matching

We compared patients with high blood glucose variability (CV above the median of 23.6%) versus low variability (CV at or below the median) in the full cohort. Before matching, the high-variability group had worse baseline characteristics (higher SOFA and disease severity scores; 28-day mortality 31.4% vs 16.2%). After 1:1 nearest-neighbour matching on the logit propensity score without replacement and with a caliper of 0.1 SD, 2,215 pairs were created and every covariate was well balanced (largest absolute SMD 0.059; Table 7); 930 of the 3,145 high-variability patients (29.6%) had no eligible control inside the caliper and were not matched. In the matched sample the difference in 28-day mortality was attenuated but remained statistically significant (26.5% vs 21.0%; log-rank $P = 1.3 \times 10^{-5}$), and Cox regression with a pair-clustered robust variance gave a hazard ratio of 1.31 (95% CI 1.17 to 1.47, $P = 4.2 \times 10^{-6}$). The estimate was not materially dependent on the matching algorithm: a wider caliper of 0.2 SD produced 2,336 pairs and a hazard ratio of 1.40 (95% CI 1.25 to 1.58), and matching with replacement produced 3,143 matched pairs and a hazard ratio of 1.10 (95% CI 1.00 to 1.20, $P = 0.048$). A large part of the crude association between glucose variability and mortality is explained by baseline illness severity. An excess hazard of about 30% persists after measured confounders are balanced, so the strong adjusted associations from the multivariable and time-varying Cox models should be read with this caveat in mind. Kaplan-Meier curves before and after matching are shown in Figure 10.

Table 7. Propensity score matching balance assessment (high vs low glycemic variability, full cohort)

Covariate	SMD (before)	SMD (after)
Age	0.090	-0.025
Sex (male)	-0.158	-0.005
Diabetes	0.509	0.059
SOFA	0.532	0.044
APS III	0.527	0.026
SAPS II	0.422	0.020
Charlson	0.373	0.028

SMD = standardized mean difference; absolute SMD < 0.1 indicates adequate balance. Matching was 1:1 nearest-neighbour on the logit of the propensity score, without replacement, caliper 0.1 SD; 2,215 pairs were formed from 3,145 eligible high-variability patients. Significance tests after matching are not reported because they are not informative for judging balance.

Diabetes stratified analysis

We repeated the main analyses separately for patients with (n=1,987) and without (n=4,303) diabetes. Blood glucose variability measures were linked to death in both groups, but some differences were found (Table 8, Figure 11). The link between TIR and death was stronger in non-diabetic patients (interaction $P = 0.0002$), suggesting that non-diabetic patients may be more sensitive to blood glucose swings. SHR was associated with death in non-diabetic patients (OR 1.40 per 1-unit increase, $P = 0.030$) but not significantly in diabetic patients (OR 1.16, $P = 0.35$), and the interaction was not significant.

Table 8. Diabetes stratified analysis of blood glucose measures and 28-day death

Variable	Non-DM OR	Non-DM P	DM OR	DM P	Interaction P
Mean glucose	1.014	<0.001	1.012	<0.001	0.28
CV	1.026	<0.001	1.015	0.005	0.15
MAGE	1.002	<0.001	1.004	0.001	0.21
TIR	0.964	<0.001	0.976	<0.001	0.0002
SHR	1.40	0.030	1.16	0.35	-

Adjusted for age, sex, and SOFA score (diabetes excluded as stratification variable). OR = odds ratio.

Sensitivity analyses

We performed several sensitivity analyses to test the robustness of our findings.

Proportional hazards assumption: Schoenfeld residual tests showed departures from proportionality that were driven mainly by the covariates rather than by the glycemic exposures. Age violated the assumption in every model (P from 1.8×10^{-8} to 9.0×10^{-12}), as did SOFA score (P from 1.5×10^{-6} to 4.1×10^{-6}) and diabetes status (P 0.010

to 0.039). Among the exposures, TIR ($P = 3.9 \times 10^{-5}$), SD ($P = 0.003$) and the trajectory-group indicator ($P = 7.2 \times 10^{-5}$) showed evidence of time-varying effects; CV did not ($P = 0.058$). We relied on time-varying Cox models, on age-stratified Cox models and on a landmark analysis (Supplementary file 3, Table S9; Figure 12).

Age-stratified Cox models: Because the proportional hazards assumption was violated for age in every model, we repeated all principal Cox models stratified on age category (<65 years, n=2,558; 65–79 years, n=2,471; ≥80 years, n=1,261), which imposes no proportional hazards assumption for age. Estimates were essentially unchanged (Figure 12): the hazard ratio per 10% increase in CV was 1.196 (95% CI 1.152 to 1.241) stratified versus 1.197 (1.153 to 1.242) unstratified, and the stratum-specific estimates were 1.185 (1.126 to 1.247) below 65 years, 1.227 (1.134 to 1.328) at 65 to 79 years and 1.185 (1.097 to 1.280) at 80 years or older; for mean glucose 1.009 (1.008 to 1.009) versus 1.009 (1.008 to 1.010); and for TIR 0.976 (0.974 to 0.978) in both. The trajectory gradient was likewise preserved (G2 1.55, G3 2.36, G4 3.49, G5 2.20 stratified, versus 1.55, 2.29, 3.48 and 2.11 unstratified). Stratifying simultaneously on age and SOFA category gave a CV hazard ratio of 1.211 (95% CI 1.168 to 1.256). Non-proportionality in SOFA persisted after stratification but did not materially change any exposure estimate.

Expanded covariate models: Adding APS III and the Charlson comorbidity index to the adjustment set did not materially change the results. CV remained significant (adjusted OR 1.017 per 1%, 95% CI 1.011 to 1.023, $P < 0.0001$), TIR remained protective (adjusted OR 0.973 per 1%, 95% CI 0.970 to 0.976, $P < 0.0001$), and the very high-stable trajectory group (G4) retained its strong association with death (adjusted OR 3.80, 95% CI 2.47 to 5.85, $P < 0.0001$), against 4.28 (2.81 to 6.54) under the base adjustment. The Cox estimate for CV per 10% was 1.143 (1.101 to 1.187). Mechanical ventilation was not recorded in the extracted data set and could not be included (Supplementary file 3, Table S9).

Multiple imputation for SHR: Because HbA1c data were available for only 26% of patients, we imputed the log-transformed SHR in 10 completed data sets (m = 10) from age, sex, diabetes, SOFA score, APS III, Charlson index, mean glucose, CV and

measurement count, and pooled the results with Rubin's rules. All SHR effect estimates in this manuscript are reported per 0.1-unit increase unless otherwise stated. After pooling with Rubin's rules, SHR remained a significant predictor (pooled OR 1.040 per 0.1-unit increase, equivalent to 1.48 per 1-unit increase, 95% CI 1.022 to 1.058, $P < 0.0001$), with a fraction of missing information of 0.63. The pooled estimate was close to the complete-case result (OR 1.026 per 0.1, $P = 0.013$), and the corresponding pooled Cox estimate was an HR of 1.024 per 0.1 (95% CI 1.018 to 1.030). HbA1c itself was the missing variable, so imputation from the remaining covariates cannot recover unmeasured HbA1c; the complete-case estimate is therefore reported as the primary result, the imputed analysis is exploratory, and residual selection bias cannot be excluded (Supplementary file 3, Table S9).

Bonferroni correction: After Bonferroni correction for 6 primary variability measures (adjusted α=0.0083), all measures except glucose range remained statistically significant (Table 2).

Propensity matching on trajectory group: In a second matching analysis, the 156 patients in the very high-stable group were matched 1:1 to patients in all other classes. Because 83.3% of this group had diabetes compared with 30.5% of the remaining classes, matching was performed exactly on diabetes status and then on the propensity score estimated from age, sex, diabetes and SOFA score. All 156 pairs were formed and balance was good (largest absolute SMD 0.018; diabetes SMD 0.000, SOFA 0.002). Matched 28-day mortality was 51.3% versus 28.8% (log-rank $P = 1.0 \times 10^{-4}$; HR 2.21, 95% CI 1.52 to 3.19, $P = 2.8 \times 10^{-5}$), indicating that the excess risk associated with this trajectory is not explained by measured baseline severity.

Measurement frequency: The number of glucose measurements per patient correlated with glycemic variability (Spearman $\rho = 0.28$). Additional adjustment for measurement count strengthened rather than weakened the association between CV and mortality (OR per 10% increase 1.30, 95% CI 1.23 to 1.38, versus 1.25, 1.18 to 1.32, without this adjustment), indicating that the finding is not an artefact of surveillance intensity (Supplementary file 3, Table S9).

Landmark analysis: Because Schoenfeld residuals indicated departure from proportional hazards, we repeated the analysis in the full cohort restricted to the 5,454 patients still alive at day 7. The association between CV and mortality persisted, although attenuated (HR per 10% increase 1.087, 95% CI 1.016 to 1.163, $P = 0.016$, versus 1.197, 95% CI 1.153 to 1.242, in the full cohort), indicating that part but not all of the association is driven by early deaths (Supplementary file 3, Table S9).

Alternative trajectory model (K5_D2): To test sensitivity to polynomial degree, we refitted the trajectory model with quadratic (degree-2) rather than cubic trajectories (log-likelihood −245,689.6, BIC 491,552.3). Class assignments agreed closely with the cubic solution (label-matched Cohen's κ=0.964), the mortality gradient was preserved (13.2% to 50.7%), and the highest-mortality class retained a strong association with death (adjusted OR 4.00, 95% CI 2.61 to 6.12, $P < 0.0001$) (Supplementary file 3, Table S9).

Survivor bias in the trajectory cohort: The GBTM analysis required at least 3 days of blood glucose data, which excluded 533 of 6,290 patients (8.5%). These excluded patients had markedly higher mortality: overall cohort deaths were 1,494 (23.8%), compared with 1,073 deaths (18.6%) in the trajectory cohort, implying that 421 of the 533 excluded patients (79.0%) died within 28 days. In other words, patients excluded from trajectory modeling were predominantly early deaths with the most severe illness. This survivor (immortal-time) bias systematically removes early deaths from trajectory classification and biases the trajectory–mortality association toward the null; the true mortality gradient across trajectory groups is likely underestimated, and the reported risk of the high-glucose trajectories (G4/G5) should be considered a lower bound.

Incremental prognostic value of blood glucose measures

We asked whether glycemic information improves risk prediction beyond routine severity assessment (Table 9, Figure 13). A reference model containing age, sex, diabetes status and the SOFA score reached an AUC of 0.802 (95% CI 0.790 to 0.815). Adding the glucose CV raised the AUC only marginally (0.807, 95% CI 0.794 to 0.819; ΔAUC 0.005, 95% CI 0.001 to 0.008; $P = 0.004$ by the DeLong test), and adding MAGE produced a similar small change (0.807, 95% CI 0.794 to 0.819; ΔAUC 0.004, 95% CI 0.003 to 0.006; $P < 0.0001$). Because a ΔAUC of this size can reach significance while

remaining clinically trivial, we also examined the integrated discrimination improvement, which requires the average predicted risk to shift in the correct direction: it was 0.009 for CV (P = 0.31) and 0.004 for MAGE (P = 0.68), neither significant. TIR performed distinctly better. Adding it raised the AUC to 0.832 (95% CI 0.821 to 0.844; ΔAUC 0.030, 95% CI 0.023 to 0.037; $P < 0.0001$), with a continuous NRI of 0.585 and an IDI of 0.056 (both $P < 0.0001$). Adding all three measures together was no better than TIR alone: the AUC was 0.8320 against 0.8323 for TIR alone (ΔAUC −0.0003), and although a likelihood ratio test of the two additional coefficients was marginally significant (χ^2 = 7.13, 2 df, P = 0.028), discrimination did not improve.

To remove any doubt that these gains merely restate illness severity, we repeated the analysis against a fuller severity model containing APS III, SAPS II and the Charlson comorbidity index, which by itself reached an AUC of 0.829 (95% CI 0.818 to 0.840). Adding CV, TIR and MAGE to this model still improved discrimination (AUC 0.848, 95% CI 0.838 to 0.858; ΔAUC 0.019, 95% CI 0.015 to 0.025; $P < 0.0001$), with a continuous NRI of 0.546 and an IDI of 0.041 (both $P < 0.0001$). The gains were reflected in calibration and in clinical usefulness: the Brier score fell from 0.133 to 0.126, the bootstrap-validated calibration slope was 0.993 with an intercept of −0.004, and decision curve analysis showed a higher net benefit than the severity-only model across threshold probabilities from about 15% to 55% (Figure 13).

Finally, in the trajectory cohort the five trajectory groups added to the reference model improved discrimination from 0.799 to 0.822 (ΔAUC 0.023, 95% CI 0.016 to 0.029; $P < 0.0001$; continuous NRI 0.547; IDI 0.024, $P = 0.017$), indicating that the trajectory phenotype carries information that age, sex, diabetes and the SOFA score do not. The optimism-corrected AUC differed from the apparent value by no more than 0.001 in every model, so these estimates are not appreciably over-fitted.

Table 9. Incremental prognostic value of blood glucose measures for 28-day death

Model	AUC (95% CI)	Optimism-corrected AUC	Brier score
A. Discrimination of nested models (n = 6,290)			

Base: age, sex, diabetes, SOFA	0.802 (0.790-0.815)	0.801	0.140
Base + glucose CV	0.807 (0.794-0.819)	0.806	0.138
Base + MAGE	0.807 (0.794-0.819)	0.805	0.139
Base + TIR 70-180	0.832 (0.821-0.844)	0.831	0.130
Base + CV + TIR + MAGE	0.832 (0.821-0.843)	0.831	0.130
Full severity: base + APS III + SAPS II + Charlson	0.829 (0.818-0.840)	0.827	0.133
Full severity + CV + TIR + MAGE	0.848 (0.838-0.858)	0.847	0.126
Base + insulin exposure (days 0-14)	0.810 (0.798-0.823)	0.809	0.138
Full severity + CV + TIR + MAGE + insulin	0.855 (0.845-0.866)	0.854	0.123
B. Added predictive ability relative to the reference model			
Comparison	ΔAUC (95% CI)	P (DeLong)	Continuous NRI
Glucose CV added to base model	0.005 (0.001-0.008)	0.004	0.269
MAGE added to base model	0.004 (0.003-0.006)	<0.0001	0.324
TIR added to base model	0.030 (0.023-0.037)	<0.0001	0.585
CV + TIR + MAGE added to base model	0.030 (0.023-0.037)	<0.0001	0.576
CV + TIR + MAGE added to full severity model	0.019 (0.015-0.025)	<0.0001	0.546
Trajectory group added to base model (n = 5,757)	0.023 (0.016-0.029)	<0.0001	0.547
Insulin exposure added to base model	0.008 (0.005-0.012)	<0.0001	0.185
Glucose CV added to base model with insulin	0.006 (0.003-0.010)	0.0002	0.316
TIR added to base model with insulin	0.034 (0.027-0.040)	<0.0001	0.621
CV + TIR + MAGE added to base model with insulin	0.034 (0.027-0.041)	<0.0001	0.626

Panel A shows apparent and optimism-corrected discrimination (200 bootstrap replications) for nested logistic models; 95% confidence intervals for the AUC come from 2,000 bootstrap resamples. Panel B shows the change in AUC, tested with the DeLong test for correlated curves, and the continuous net reclassification improvement (NRI). The integrated discrimination improvement (IDI) was 0.009 (P = 0.31) for CV, 0.004 (P = 0.68) for MAGE, 0.056 (P < 0.0001) for TIR, 0.057 (P < 0.0001) for CV + TIR + MAGE added to the base model, 0.041 (P < 0.0001) for CV + TIR + MAGE added to the full severity model, and 0.024 (P = 0.017) for the trajectory groups. The category-based NRI at the observed 28-day mortality of 23.8% was 0.016 (P = 0.020) for CV, 0.018 (P = 0.0001) for MAGE, 0.045 (P < 0.0001) for TIR, 0.045 (P < 0.0001) for CV + TIR + MAGE added to the base model, 0.024 (P = 0.017) for CV + TIR + MAGE added to the full severity model, and 0.039 (P = 0.0003) for the trajectory groups. All models were estimated by unpenalised maximum likelihood. Bootstrap-validated calibration (200 replications) was good throughout: the calibration slope was 0.995 for the base model, 0.995 for base plus TIR, 0.993 for base plus CV, TIR and MAGE and 0.993 for the full severity model with all three measures, with calibration intercepts between −0.004 and −0.002 and optimism-corrected Brier scores of 0.140, 0.130, 0.130 and 0.126 respectively. A likelihood ratio test of the two coefficients added by CV and MAGE to a model that already contained TIR gave χ^2 = 7.13 (2 df, P = 0.028) while the AUC fell from 0.8323 to 0.8320. AUC = area under the receiver operating characteristic curve; CV = coefficient of variation; TIR = time in range; MAGE = mean amplitude of glycemic excursion; IDI = integrated discrimination improvement.

Insulin exposure and adjustment for glucose-lowering therapy

Insulin was given to 5,302 of the 6,290 patients (84.3%) during days 0 to 14. Exposure was almost universal in patients with diabetes (97.7%) but also common in patients without it (78.1%), consistent with treatment of stress hyperglycemia. It rose steeply across the trajectory groups, from 80.6% in G1 to 91.5% in G2, 97.2% in G3, 99.3% in G5 and 100% in G4 (χ^2 = 224.3, 4 df, P < 0.0001), as did the mean daily dose, from 5.9 units in G1 to 26.7 units in G4 (Table 1, Table 4). In univariable comparisons more patients who survived received insulin than patients who died, and they received it for

more days, but this comparison is uninformative because it reflects opportunity for exposure rather than any effect of treatment.

Adjustment for insulin did not attenuate any of the associations, and slightly strengthened all of them. In the whole cohort the odds ratio per 1% increase in CV was 1.022 (95% CI 1.016 to 1.028) before adjustment and 1.026 (1.020 to 1.032) after it; for TIR it was 0.970 (0.966 to 0.973) and 0.967 (0.964 to 0.970); for MAGE 1.003 (1.001 to 1.004) and 1.003 (1.002 to 1.004); and for mean glucose 1.013 (1.011 to 1.015) and 1.015 (1.013 to 1.017) (Supplementary file 1, Table S1). The same pattern held for the trajectory groups, whose adjusted odds ratios rose rather than fell when insulin was added. Insulin itself was associated with substantially lower odds of death, with odds ratios between 0.33 and 0.40, a magnitude too large to be causal and characteristic of confounding by indication together with the immortal time implicit in a cumulative 14-day exposure.

The temporal analyses address that bias and reach the same conclusion. In the time-varying Cox model, in which exposure is updated each day, same-day insulin was associated with a hazard ratio of 0.73 (95% CI 0.62 to 0.87, $P = 0.0003$) while daily CV remained strongly associated with death (HR 1.016 per 1%, 95% CI 1.011 to 1.021, $P < 0.0001$) and daily mean glucose likewise (HR 1.005 per 1 mg/dL, 95% CI 1.004 to 1.006, $P < 0.0001$) (Table 6, Model 4). In the seven-day landmark analysis, restricted to the 5,454 patients still alive and under observation on day 7, adding insulin again moved the estimates away from the null: the odds ratio per 1% increase in CV was 1.009 (1.001 to 1.017) before and 1.013 (1.005 to 1.021) after adjustment, and for TIR 0.979 (0.974 to 0.983) and 0.976 (0.972 to 0.980). Insulin exposure in the first seven days had an odds ratio of 0.28 to 0.33, again indicating confounding by indication rather than a treatment effect.

The dissociation between association and incremental information survived adjustment as well (Table 9). Insulin exposure itself added 0.008 to the area under the curve of the base model (95% CI 0.005 to 0.012, $P < 0.0001$). When insulin is already in the reference model, CV adds 0.006 (0.003 to 0.010, $P = 0.0002$) whereas TIR adds 0.034 (0.027 to 0.040, $P < 0.0001$), a six-fold difference that is the same as the difference observed without insulin, and adding CV, TIR and MAGE together to a model containing insulin

gives 0.034, no more than TIR alone. Finally, the CV association was stronger in patients who received insulin than in those who did not (interaction $P = 0.0033$), whereas TIR behaved identically in both strata (interaction $P = 0.22$) (Supplementary file 1, Table S2). Taken together, these analyses make it unlikely that glucose-lowering therapy accounts for the prognostic signal of either glycemic variability or reduced time in range.

A marginal structural model was then fitted to separate the effect of treatment from the severity that prompts it. Exposure and confounders were measured on the same ICU day and death during the following day was the outcome, giving 34,824 person-days, 6,285 patients and 936 deaths within the observation window (Supplementary file 1, Table S4). For insulin the stabilised weights were close to 1, because exposure is highly persistent from one day to the next, and the weighted estimate was consequently almost identical to the unweighted one (odds ratio 0.53 per exposed day, 95% CI 0.44 to 0.63, against 0.58, 0.49 to 0.67 without weighting). Weighting therefore does not turn this parameter into a treatment effect: its direction and its magnitude both indicate confounding by indication, which no observational method can remove and which we do not read as benefit. For a day with glucose outside 70–180 mg/dL the weights were more variable and weighting moved the estimate part of the way towards the null, from 1.65 (1.43 to 1.91) to 1.35 (1.08 to 1.69) per day. A substantial share of the crude day-to-day association therefore reflects time-varying treatment and severity; the association nevertheless remained positive and statistically significant, and the incremental discrimination analyses, which do not depend on the exposure-outcome association, were unaffected.

External validation in the eICU Collaborative Research Database

Of 200,859 ICU admissions in eICU-CRD, 138,942 were the first ICU stay of a patient, 96,567 lasted at least 24 hours, 27,410 met Sepsis-3 criteria, and 15,545 of these had glucose recorded on at least three days and entered the trajectory analysis (Table 10). Compared with MIMIC-IV, these patients were slightly younger (63.8 vs 66.8 years), more often female (47.2% vs 38.7%) and less often diabetic (24.0% vs 31.9%), whereas overall glucose control was similar (mean glucose 142.9 vs 140.8 mg/dL; CV 24.2% vs 25.7%; TIR 80.1% vs 83.1%). Hospital mortality was 18.6% and ICU mortality 11.2%.

The same five-class cubic specification was recovered. From five independent random starts the EM algorithm converged to an identical log-likelihood (−509,440.1) and the average posterior probability of assignment was 0.882. Exactly as in MIMIC-IV, the Bayesian Information Criterion continued to fall as classes were added, so the number of classes was fixed a priori at five to preserve comparability with the development cohort. The five groups reproduced the clinical spectrum of the original cohort: a large low-stable group (G1, 64.2%, mean glucose 119.5 mg/dL), a small intermediate group rising over the stay (G2, 3.6%), a mid-range group (G3, 23.6%), a high group peaking late (G4, 5.1%) and a very high group falling steeply (G5, 3.5%). Hospital mortality was lowest in G1 (15.8%) and higher in each of the other groups (31.3%, 23.5%, 21.5% and 20.6% for G2 to G5; $P < 0.0001$). After adjustment for age, sex, diabetes, SOFA and APACHE IV, every group retained a significantly higher odds of death than G1 (G2 adjusted OR 2.47, 95% CI 2.01 to 3.02; G3 OR 1.54, 1.39 to 1.71; G4 OR 1.60, 1.32 to 1.95; G5 OR 1.39, 1.10 to 1.77). The gradient therefore reproduced, although the group carrying the highest risk differed: in MIMIC-IV it was the persistently hyperglycemic class, whereas in eICU-CRD it was the group with the steepest upward excursion.

When the MIMIC-IV parameters were transported unchanged to eICU-CRD, the average posterior probability of assignment was 0.826 and the imported class proportions closely followed those of the development cohort (56.8%, 20.2%, 11.9%, 4.2% and 6.9% versus 64.9%, 19.7%, 8.0%, 2.7% and 4.7%). The five transported mean curves lay almost on top of the MIMIC-IV curves, with a mean root-mean-square difference of 6.6 mg/dL (range 3.6 to 12.4) and a mean shape correlation of 0.76; agreement between the refitted and the transported assignment was substantial (adjusted Rand index 0.70; Figure 14).

All variability measures showed the same direction of association as in MIMIC-IV. Per unit increase, the adjusted odds ratio for hospital mortality was 1.010 (95% CI 1.007 to 1.014) for CV, 1.004 (1.003 to 1.006) for SD, 1.002 (1.001 to 1.002) for MAGE and 1.002 (1.001 to 1.002) for range, whereas each 1% increase in TIR was associated with an odds ratio of 0.992 (0.990 to 0.994); all $P < 0.0001$, and the estimates for ICU mortality were almost identical (Table 10). Discrimination was lower in eICU-CRD than in MIMIC-IV in absolute terms, reflecting the multicentre case mix across 208 hospitals and the different outcome definition, so absolute areas under the curve are not

comparable between the two databases. The incremental pattern, however, was the same: adding CV to the severity model raised the area under the curve by only 0.0050 (95% CI 0.0026 to 0.0073), whereas TIR raised it by 0.0083 (0.0053 to 0.0119), and the trajectory group added 0.0139 (0.0095 to 0.0183), the largest increment of any single exposure. As in MIMIC-IV, TIR was the stronger incremental marker.

Insulin exposure was recorded for 18.5% of the cohort and for 19.3% of the 14,243 patients with complete covariates. Adding it to the model left the association with CV essentially unchanged (OR 1.0097, 95% CI 1.0064 to 1.0131 versus 1.0102, 1.0069 to 1.0135) and left the association with time in range likewise unchanged (0.9922, 0.9904 to 0.9940 versus 0.9920, 0.9902 to 0.9938). Insulin itself was not independently associated with death (OR 1.10, 95% CI 0.98 to 1.23; $P = 0.10$), and the increment in discrimination attributable to CV or to time in range was the same whether or not insulin was already in the model (0.0045 versus 0.0050 for CV, and 0.0079 versus 0.0084 for time in range), whereas insulin itself added 0.0007 ($P = 0.20$). The variability signal in this cohort is therefore not explained by glucose-lowering therapy (Supplementary file 1, Table S3).

Table 10. External validation in the eICU Collaborative Research Database

Variable	**MIMIC-IV (n = 5,757)**	**eICU-CRD (n = 15,545)**
A. Cohort characteristics		
Age, years, mean (SD)	66.8 (15.5)	63.8 (15.7)
Male, n (%)	3,530 (61.3)	8,206 (52.8)
Diabetes, n (%)	1,837 (31.9)	3,731 (24.0)
SOFA score, mean (SD)	9.1 (3.6)	7.7 (2.6)
Severity score, mean (SD)	APS III 54.7 (18.1)	APACHE IV 72.0 (27.4)
Mean glucose, mg/dL	140.8	142.9
CV, %	25.7	24.2
TIR 70-180 mg/dL, %	83.1	80.1
MAGE, mg/dL	62.4	63.6
Any glucose < 70 mg/dL, n (%)	1,434 (24.9)	2,456 (15.8)
Mortality, n (%)	28-day 1,071 (18.6); hospital 979 (17.0)	hospital 2,895 (18.6); ICU 1,747 (11.2)
B. Trajectory groups refitted de novo in eICU-CRD		

Group	n (%); mean glucose; CV; TIR; hospital mortality	Adjusted OR (95% CI)
G1 low-stable	9,973 (64.2); 119.5; 21.1; 93.4; 15.8%	1.00 (reference)
G2 intermediate rising	559 (3.6); 188.5; 33.5; 47.8; 31.3%	2.47 (2.01-3.02)
G3 mid-range	3,672 (23.6); 167.5; 26.7; 66.6; 23.5%	1.54 (1.39-1.71)
G4 high, late peak	796 (5.1); 226.9; 30.8; 28.0; 21.5%	1.60 (1.32-1.95)
G5 very high, falling	545 (3.5); 235.9; 44.1; 37.6; 20.6%	1.39 (1.10-1.77)
C. Glycemic variability and hospital mortality in eICU-CRD (adjusted)		
Variable	Unit	Adjusted OR (95% CI); P
CV	per 1%	1.010 (1.007-1.014); < 0.0001
SD	per 1 mg/dL	1.004 (1.003-1.006); < 0.0001
TIR 70-180	per 1%	0.992 (0.990-0.994); < 0.0001
MAGE	per 1 mg/dL	1.002 (1.001-1.002); < 0.0001
Range	per 1 mg/dL	1.002 (1.001-1.002); < 0.0001
Mean glucose	per 1 mg/dL	1.004 (1.003-1.005); < 0.0001
D. Incremental discrimination for hospital mortality in eICU-CRD		
Model	AUC	Change in AUC (95% CI); P
Base: age, sex, diabetes, SOFA	0.643	—
Full severity (+ APACHE IV)	0.679	—
Full severity + CV	0.684	+0.0050 (0.0026-0.0073); < 0.0001
Full severity + TIR	0.688	+0.0083 (0.0053-0.0119); < 0.0001
Full severity + CV + TIR + MAGE	0.689	+0.0101 (0.0065-0.0140); < 0.0001
Full severity + trajectory group	0.693	+0.0139 (0.0095-0.0183); < 0.0001

Variables are shown as mean (SD) or n (%). Adjusted odds ratios are from logistic regression including age, sex, diabetes, SOFA and APACHE IV. AUC = area under the receiver operating characteristic curve; CV = coefficient of variation; MAGE = mean

amplitude of glycemic excursion; TIR = time in range; APS III = Acute Physiology Score III; APACHE IV = Acute Physiology and Chronic Health Evaluation IV.

Discussion

In this study of 6,290 patients with sepsis, both the glucose level and its swing over time were linked to death. We identified five glucose patterns during the ICU stay, with death rates from 13% to 51%. Multivariable regression and time-varying Cox models supported these findings. After propensity score matching on baseline severity, the variability association fell by about two thirds but stayed significant (HR 1.31, 95% CI 1.17 to 1.47). The very high-stable trajectory kept a more than two-fold hazard after matching (HR 2.21, 95% CI 1.52 to 3.19). Residual confounding calls for caution about the size of the variability effect; the trajectory gradient is less affected.

Comparison with previous studies

Our results agree with earlier work linking high glucose variability to worse outcomes in critically ill patients [14, 15]. Eslami et al. found that several variability measures predicted death in ICU patients, although they noted that most studies did not check model assumptions [14]. Ali et al. reported the same in a mixed ICU population that included patients with sepsis [15]. Krinsley found a strong independent association between glycemic variability and mortality, strongest in the normoglycemic range [16], and Egi et al. reported the same pattern in 7,049 patients [17]. Hermanides et al. confirmed the association in a separate cohort [18]. Our study adds two things: the association holds in a sepsis-specific cohort, and trajectory modeling captures the day-to-day pattern that a single summary number misses.

Several recent studies have applied GBTM to blood glucose trajectories in sepsis. Yang et al. identified three glucose trajectory groups in 3,869 patients with sepsis and ARDS [30], and Wang et al. used a similar three-group model in 8,923 patients with sepsis [31]. Ma et al. used K-means clustering in 3,655 patients and identified five trajectory clusters, replicated externally in 10,874 MIMIC-IV patients [32], supporting a five-group solution. These studies focused on classification alone. Our study differs in three ways: (1) we identified five trajectory groups, which revealed subtypes such as the very high-stable

(2.7%, 51.3% mortality) and high-start declining (4.7%) patterns that three-group models cannot separate; (2) we combined trajectory analysis with SHR, several variability measures, restricted cubic splines, time-varying Cox regression and propensity score matching in one cohort; and (3) we formally tested interaction by diabetes, showing that patients without diabetes are more vulnerable to glucose variability (interaction P = 0.0002 for TIR).

The SHR finding fits with earlier work. Roberts et al. showed that SHR predicted death better than admission glucose in patients with critical illness [46]. A dose-response meta-analysis in critically ill patients reached the same conclusion [23], and Li et al. reported an independent association in 13,199 patients with sepsis [24]. Wang et al. combined SHR with glucose variability and machine learning (XGBoost, random forest) in 4,838 critically ill patients and found that SHR-based models outperformed conventional scores [47]. Our study extends these findings: SHR remained an independent predictor after adjustment for disease severity, and the spline analysis showed a J-shaped relation with a risk inflection near SHR 1.5. We chose trajectory phenotyping over a machine-learning model because it is easier to interpret and does not require model deployment at the bedside.

Association versus incremental prognostic information

The most instructive finding of this study is the dissociation between how strongly a glycemic measure is associated with death and how much it adds to risk prediction. Glucose CV was among the strongest exposures we examined: the adjusted odds of death rose by 2.2% per 1% increase in CV, the association survived propensity matching, adjustment for measurement frequency and a day-7 landmark analysis, and the hazard ratio remained 1.20 per 10% in time-varying models. Yet adding CV to a model containing age, sex, diabetes and SOFA raised the AUC by only 0.005 (95% CI 0.001 to 0.008; P = 0.004), one sixth of the increment achieved by time in range, and left the integrated discrimination improvement at 0.009 (P = 0.31), which is not statistically significant. Time in range behaved differently: it raised the AUC by 0.030 (P < 0.0001) and reclassified more than twice as many patients as CV did (continuous NRI 0.585 versus 0.269). The contrast is sharpened by the nested comparison: a model containing

TIR alone reached an AUC of 0.8323 and a model containing CV, TIR and MAGE together reached 0.8320 (Table 9), so once TIR was known, variability added nothing measurable to discrimination, even though the likelihood ratio test for the two additional coefficients was marginally significant (P = 0.028). We therefore describe the contribution of variability to discrimination as detectable but negligible, and that of time in range as substantial.

We interpret this as evidence that the prognostic content of glycemic variability is largely shared with illness severity. Variability is generated in part by the processes that determine severity: organ failure, interrupted enteral feeding, vasopressor and corticosteroid exposure, renal replacement therapy, and the intensity of monitoring itself, which in this cohort was correlated with the number of measurements. All of these are already represented in SOFA, APS III and SAPS II. Time in range, by contrast, summarises the net duration of exposure to a defined glucose band and therefore retains information about glycaemic control that severity scores do not contain. Trajectory membership added 0.023 to the AUC (95% CI 0.016 to 0.029; P < 0.0001), intermediate between the two, which is consistent with trajectories encoding both level and persistence.

This dissociation has two consequences. First, it offers a parsimonious explanation for the conflicting literature on glycemic variability: studies that adjusted for a limited covariate set, or that reported association without discrimination, would be expected to find variability prognostic, whereas comparisons against a full severity model, as performed here, find little residual information. Second, it argues against adding variability indices to risk prediction models in sepsis, and against treating variability as an intervention target on the strength of observational association alone. Neither conclusion denies that wide glucose excursions may be biologically harmful; oxidative stress and endothelial injury remain plausible mechanisms. The claim is narrower: at the level of group-level risk prediction, variability is a marker of how ill a patient is rather than an additional piece of prognostic information, and only a trial in which variability is modified independently of severity can establish causality.

Blood glucose trajectory patterns

The five glucose patterns add detail that three-group models cannot show. The most important finding is the very high-stable group (Group 4). It included only 2.7% of patients, and half of them died. Glucose stayed above 205 mg/dL throughout the ICU stay (adjusted OR 4.28, 95% CI 2.81 to 6.54), and the excess risk persisted after propensity matching on age, sex, diabetes and SOFA score (HR 2.21, 95% CI 1.52 to 3.19). This phenotype deserves testing in future interventional studies.

The high-start declining group (Group 5) is also worth attention. Glucose fell from 264 mg/dL at admission to 139 mg/dL by day 14. The 28-day death rate was still 33.3%. Early hyperglycemia may cause lasting harm even when glucose is brought down later.

How much do these measures add?

The incremental analyses put a bound on the clinical value of what we report. Glycemic information improved discrimination beyond a full severity model, but the gain was modest (ΔAUC about 0.02) and came almost entirely from TIR rather than from the CV; the CV raised the AUC by 0.005 (P = 0.004) but did not significantly improve the integrated discrimination improvement (0.009; P = 0.31) despite its strong adjusted association with death. This dissociation is informative. The CV correlates with illness severity, and once SOFA, APS III, SAPS II and the Charlson index are in the model most of its prognostic signal has already been captured, which matches the attenuation we observed after propensity matching. TIR, by contrast, contributed information that severity scores did not contain. The practical implication is that a single well-defined measure such as TIR may be more useful for risk models than variability indices, and that strong adjusted associations for variability should be read alongside a discrimination analysis rather than instead of it. The trajectory phenotype itself also improved discrimination (ΔAUC 0.023), so the day-to-day pattern is not fully reducible to its summary measures. We are not aware of another glucose-trajectory study in sepsis that has tested incremental discrimination and reclassification against a full severity model, and this is, in our view, the question that determines whether trajectory phenotyping is worth translating to the bedside.

Possible mechanisms

Several mechanisms may explain the link between glucose variability and death. Glucose swings raise oxidative stress more than a constant high level does [48, 49]. Acute hyperglycemia weakens immune function [50, 51] and changes vascular tone [52]. Both high and low glucose can add to the inflammatory burden already present in sepsis [9]. Oscillating glucose also impairs the antioxidant response of endothelial cells [53]. Whether these mechanisms translate into harm in patients is still debated [21]. The usual glucose level modifies the effect: patients with diabetes seem less sensitive to acute swings than those without [54, 55]. This fits with our finding that variability matters more in patients without diabetes (interaction P = 0.0002 for TIR).

Clinical implications and the NICE-SUGAR context

Our findings should be read against the NICE-SUGAR trial, in which intensive glucose control (target 81-108 mg/dL) increased mortality compared with conventional control (target below 180 mg/dL) [56]. The Leuven trials reported lower mortality with intensive insulin therapy in surgical and medical ICU patients [10, 57], a result later trials did not reproduce. Our results do not support tighter glucose control. They show that sustained hyperglycemia and wide glucose swings are both linked to poor outcomes, and that the very high-stable group (glucose above 220 mg/dL) is a high-risk phenotype worth studying further. Hypoglycemia is also associated with death in the ICU [58], and the risk it confers is itself modified by pre-morbid glycemic control: in a three-centre study of 3,084 critically ill patients, hypoglycaemia was independently associated with hospital mortality only among those whose pre-admission HbA1c was 8.0% or higher (interaction P = 0.009) [59]. Pushing mean glucose down therefore carries risk, and this is the conclusion of the most recent synthesis of the trial evidence: a network meta-analysis of 37 trials and 27,541 patients, published in this journal, ranked a target of 144–180 mg/dL highest for both survival and freedom from hypoglycaemia, whereas 110–144 mg/dL ranked highest only for infection and acute kidney injury [60]. A target in the middle of the range we observed (mean glucose about 125 mg/dL in the lowest-risk trajectory) is therefore a defensible clinical anchor, and our data add that the aim should be to keep glucose inside 70–180 mg/dL for as much of the ICU stay as possible while avoiding

both sustained hyperglycaemia above 200 mg/dL and repeated hypoglycaemia. Whether deliberately reducing the size of the excursions, rather than the mean, improves survival has not been tested. That question needs a randomized trial before any causal claim can be made.

External validation in a multicentre database

The central claim of this study was that glucose level, glucose variability and the shape of the glucose trajectory carry independent prognostic information in sepsis. Because that claim rested on a single-centre cohort, we tested it in eICU-CRD, an independent registry of 208 hospitals. Three findings emerged. First, the five-class structure was reproducible: refitting the model from scratch recovered a numerically identical optimum from five independent random starts, exactly as in MIMIC-IV, and the class proportions were of the same order (a dominant low-stable group of about two thirds of patients with a long tail of smaller high-glucose classes). Second, the model was transportable: when the MIMIC-IV parameters were frozen and applied to eICU-CRD patients, the reconstructed mean curves lay within a mean of 6.6 mg/dL of the original curves and the average posterior probability of assignment was 0.826, well above the conventional 0.70 threshold for adequate class separation. Third, every glycemic exposure kept its direction and its statistical significance, and the pattern of incremental discrimination was identical: TIR added roughly twice as much discrimination as CV in both databases, although neither increment was large in absolute terms.

One discrepancy deserves emphasis rather than concealment. The mortality gradient reproduced, but the group carrying the highest risk did not: in MIMIC-IV the highest mortality was seen in the persistently hyperglycemic class, whereas in eICU-CRD it was the smaller class with the steepest upward excursion. Several explanations are plausible. The two cohorts differ in case mix, in the density of glucose sampling and in glucose-management protocols, and the eICU-CRD patients were less severely ill as judged by the class-specific severity scores, which were almost flat across groups. Small high-glucose classes are also the least stable part of any trajectory solution, and their boundaries should not be treated as fixed thresholds. The practical implication is that the

ordinal position of a trajectory group, not the identity of a particular class label, carries the prognostic signal.

Generalizability across health systems.

Both databases we analysed come from the United States, and the readership of this journal practises in health systems that differ in the prevalence of diabetes, in the availability of continuous glucose monitoring, in nutrition and insulin protocols and in the structure of post-ICU follow-up. The trajectory solution was reproducible and transportable across 208 hospitals within one country, which addresses between-centre heterogeneity, but it does not establish that the same five phenotypes, or the same absolute mortality gradient, hold elsewhere. National guidance is not uniform either: the Japanese Clinical Practice Guidelines for Management of Sepsis and Septic Shock 2020 set out their own recommendations for sepsis care in Japanese practice [61], and the absolute glucose distributions we report may not apply where case mix, nutrition practice or monitoring density differ. We therefore present the class structure and the relative ordering of risk, rather than the numerical boundaries of the classes, as the transferable part of the result, and we state explicitly that prospective confirmation in other health systems is required before these phenotypes are used for triage or as the target of an intervention.

Both cohorts also allowed us to address a confounding pathway that observational studies of glucose usually leave open. Insulin is the dominant determinant of both glucose level and glucose variability in the ICU, and it is simultaneously a response to hyperglycemia and a cause of hypoglycemia, which makes it a time-varying confounder influenced by the very glucose values it is intended to correct. Recording differed between the two databases, as registries of routine care do: 84.3% of the MIMIC-IV cohort received insulin, whereas 18.5% of the eICU-CRD cohort had insulin exposure recorded, a gap that reflects how much of the subcutaneous and sliding-scale treatment each database captures rather than a difference in practice, and is one reason why absolute exposure rates should not be transported between systems. The conclusion, however, was identical in both. Adjusting for exposure left every estimate unchanged or moved it away from the null, adding variability and time in range to a model that already contained insulin

reproduced the same six-fold difference in incremental information, and treating insulin as a daily updated exposure did not diminish either signal. A marginal structural model with stabilised inverse-probability-of-treatment weights, in which exposure and confounders were updated daily and death on the following day was the outcome, likewise left the insulin estimate essentially unchanged and far below one, while moving the estimate for a day outside the target range part of the way towards the null (from 1.65 to 1.35 per day) without removing it (Supplementary file 1, Table S7). The two signals therefore respond differently to a method that addresses time-varying treatment explicitly, and neither is explained by glucose-lowering therapy. Insulin itself carried odds ratios well below one, a magnitude that no plausible treatment effect explains and that points to confounding by indication together with immortal time. What survives is the conclusion that glucose-lowering therapy is not an adequate explanation for either the variability or the time-in-range signal.

Strengths and limitations

This study has several strengths. We brought trajectory modeling, the stress hyperglycemia ratio and several variability measures together in one cohort, which allows direct comparison, and we identified five trajectory groups rather than three, which revealed subtypes such as the very high-stable group (2.7%, 51.3% mortality) that coarser models would miss. Multivariable regression, splines and time-varying Cox models gave consistent results, and we report openly that matching reduced but did not remove the variability association while the trajectory gradient persisted. Sensitivity analyses covering proportional hazards, expanded covariate models, multiple imputation, Bonferroni correction and a marginal structural model supported the main findings.

Several limitations should be noted. The study is retrospective and observational, so it describes association rather than causation; the fall in the variability–mortality association after propensity score matching (HR 1.31, 95% CI 1.17 to 1.47, from a crude HR of 2.14) shows how much of it is attributable to baseline severity, and although time-varying Cox models and the marginal structural model adjust for severity and for treatment intensity as they evolve, residual confounding cannot be excluded. The trajectory analysis required at least three days of glucose data and therefore excluded

early deaths (421 of the 533 excluded patients, 79.0%, died within 28 days); this survivor bias pushes the trajectory–mortality association towards the null, so the true risk of the high-glucose phenotypes is probably higher than we observe. HbA1c was available for 26% of patients and that subgroup was healthier (mortality 12.6% versus 23.8% overall), which indicates non-random missingness that multiple imputation cannot fully correct. We did not adjust for vasopressor use, mechanical ventilation or nutritional support, and glucose measurement frequency varies with length of stay and may itself contribute to the variability signal. Glucose was sampled intermittently rather than continuously, so MAGE computed from these samples only approximates the continuous-monitoring metric.

Three modelling choices deserve comment. The latent class model assumes that patients within a class share a common mean trajectory; mixture likelihoods are multi-modal, and the retained five-class solution is one of several local optima of similar fit, which is why the complete model grid, the R cross-validation and the re-initialisation checks are reported in Supplementary file 2 rather than the solution being treated as unique: the R fit from its own starting values reached a marginally lower log-likelihood with an equally well-separated partition, and re-initialising it at the retained solution returned it to that optimum with 99.95% of patients in the same class. BIC also continued to improve beyond five classes, and a strictly BIC-driven selection would have favoured seven. The proportional hazards assumption was violated for age in every model and for SOFA score, diabetes and some exposures; time-varying Cox models, age-stratified models and the landmark analysis gave materially identical estimates, but residual time-dependence cannot be fully excluded.

The propensity-matched estimate for variability depends on the matching algorithm (HR 1.10 to 1.40 across caliper widths and with or without replacement), and 30% of high-variability patients could not be matched within the caliper, so it applies to the region of propensity-score overlap rather than to the whole cohort. Adjustment for glucose-lowering therapy also remains imperfect: we recorded whether insulin was given but not how the dose was titrated to the preceding glucose value, and no observational method, including the marginal structural model, separates the effect of treatment from the hyperglycaemia that prompted it. The same caveat applies to corticosteroid exposure,

nutritional strategy and renal replacement therapy, none of which we could adjust for. Insulin was documented far less completely in eICU-CRD than in MIMIC-IV, so the exposure rates of the two cohorts reflect documentation as much as practice and should not be compared directly.

Finally, although the trajectory structure, its transportability and the direction of every glycemic association were reproduced in eICU-CRD, some elements remain unvalidated. The stress hyperglycemia ratio could not be examined externally because eICU-CRD contains no HbA1c, so that part of the analysis still rests on one database and on the 26% of patients with a recorded value. The external outcome was hospital and ICU mortality rather than 28-day mortality, because eICU-CRD does not record vital status after discharge. Glucose sampling was sparser in eICU-CRD (median 10 measurements per patient), which may attenuate variability metrics and explains the lower recorded frequency of hypoglycaemia, and absolute discrimination was lower there, so the two databases can be compared only in terms of the incremental value of a marker. Class proportions and the boundaries of the high-glucose classes differed between the cohorts and should not be treated as fixed thresholds; prospective evaluation is required before these phenotypes are used for risk stratification. Each sensitivity analysis returns a single estimate for one defined model and they are therefore reported together in Supplementary file 3, with the only inherently graphical comparison shown in Figure 12.

Conclusions

In patients with sepsis, blood glucose level and glycemic variability were both independently associated with 28-day mortality, but only the former carried prognostic information beyond established severity scores. Adding glucose variability to a severity-based model raised the AUC by 0.005 (95% CI 0.001 to 0.008; P = 0.004) against 0.030 for time in range and 0.023 for the glucose trajectory group, and did not improve the integrated discrimination improvement (0.009; P = 0.31); combining variability measures with time in range produced no further gain (AUC 0.8323 versus 0.8320; Table 9). This dissociation between association and discrimination indicates that the prognostic content of glycemic variability is largely captured by illness severity, which offers a parsimonious explanation for the inconsistent results of earlier variability studies. Two

implications follow. First, when glycemic information is used to augment risk prediction in sepsis, time in range is the metric to add; variability indices behave as markers of severity rather than as independent predictors. Second, the very high-stable phenotype (glucose persistently above 205 mg/dL; 51.3% mortality; HR 2.21 after propensity matching) identifies a high-risk group that merits evaluation as an enrichment criterion in future trials, but variability itself should not be assumed to be a modifiable risk factor until an intervention study shows that lowering it changes outcome independently of severity. The five-class structure was reproduced and transported in an independent database of 208 hospitals; prospective confirmation in other health systems is still required before bedside use.

List of abbreviations

APSIII: Acute Physiology Score III;
AUC: Area under the receiver operating characteristic curve;
AvePP: Average posterior probability of assignment;
BIC: Bayesian Information Criterion;
CI: Confidence interval;
CV: Coefficient of variation;
eAG: Estimated average glucose;
eICU-CRD: eICU Collaborative Research Database;
DCA: Decision curve analysis;
GBTM: Group-based trajectory modeling;
HbA1c: Glycated hemoglobin;
HR: Hazard ratio;
ICU: Intensive care unit;
IDI: Integrated discrimination improvement;
IQR: Interquartile range;
LCGA: Latent class growth analysis;
MAGE: Mean amplitude of glycemic excursion;

MIMIC-IV: Medical Information Mart for Intensive Care IV;
NRI: Net reclassification improvement;
OR: Odds ratio;
PSM: Propensity score matching;
RCS: Restricted cubic splines;
SAPSII: Simplified Acute Physiology Score II;
SD: Standard deviation;
SHR: Stress hyperglycemia ratio;
SMD: Standardized mean difference;
SOFA: Sequential Organ Failure Assessment;
TIR: Time in range

Declarations

Ethics approval and consent to participate: The MIMIC-IV database was approved by the Institutional Review Board of Beth Israel Deaconess Medical Center. The eICU Collaborative Research Database was collected under the oversight of the same Institutional Review Board; all records are de-identified and access is governed by the PhysioNet credentialed data use agreement (record ID: 78245640). For both databases the requirement for informed consent was waived. No additional ethics approval was required because the present study used only de-identified secondary data, and no individual participant can be identified.

Consent for publication: Not applicable.

Availability of data and materials: The MIMIC-IV database is publicly available at https://physionet.org/content/mimiciv/ and the eICU Collaborative Research Database at https://physionet.org/content/eicu-crd/, in both cases after completing the required training and signing a data use agreement. All extraction, analysis and figure code is available from the corresponding author on reasonable request and will be deposited in a public repository (Zenodo) with a permanent DOI upon acceptance. No new primary data

were generated in this study, and the derived tables that underlie every figure and table of this article are reported in full within the manuscript.

Competing interests: The authors declare that they have no competing interests.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Clinical trial number: Not applicable.

Acknowledgements: The authors would like to acknowledge the Massachusetts Institute of Technology Laboratory for Computational Physiology and the Beth Israel Deaconess Medical Center for creating and maintaining the MIMIC-IV database and eICU Collaborative Research Database, which made this study possible. We also acknowledge the PhysioNet team for providing open access to these databases.

Supplementary files

Supplementary file 1. Insulin exposure and time-varying confounding: exposure by vital status and by trajectory group, glucose measures before and after adjustment for insulin, incremental discrimination with insulin in the reference model, analyses stratified by insulin exposure with interaction tests, seven-day landmark analyses and the time-varying Cox model.

Supplementary file 2. Model selection for group-based trajectory modeling: all 18 fitted models with log-likelihood, BIC, average posterior probability and class sizes.

Supplementary file 3. Sensitivity analyses and proportional-hazards diagnostics.

Figures

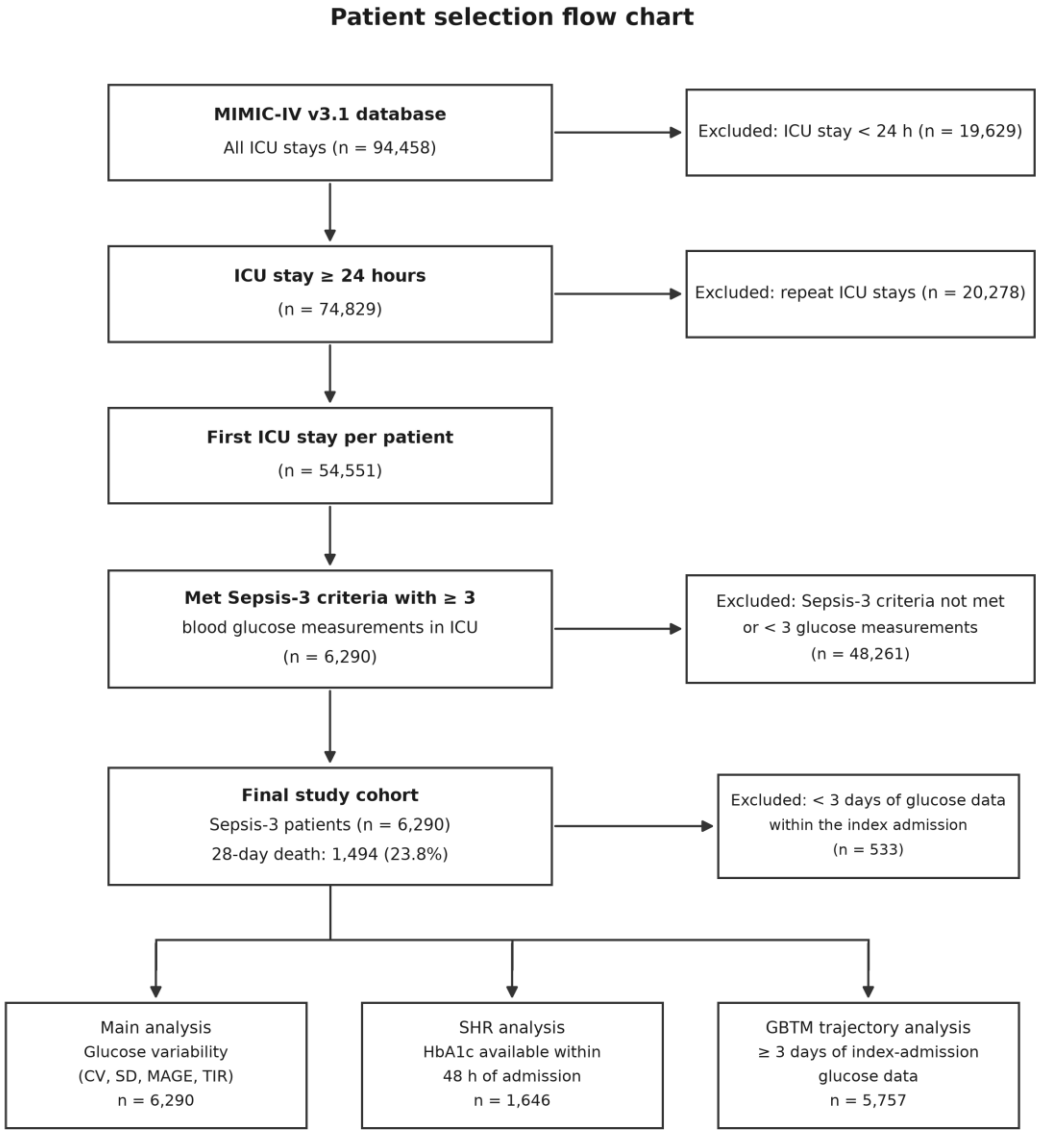


Figure 1. Patient selection flow chart. Of the 6,290 patients in the cohort, 5,757 had at least 3 days of glucose data within the index hospital admission and entered the trajectory analysis; all 5,757 received a class assignment. The external validation cohort was assembled in the eICU Collaborative Research Database with identical eligibility criteria (200,859 ICU admissions, 138,942 index stays, 96,567 stays of at least 24 hours, 27,410 meeting Sepsis-3 criteria, 20,468 with at least three glucose measurements and 15,545 with glucose recorded on at least three days; Table 10). MIMIC-IV = Medical

Information Mart for Intensive Care IV; eICU-CRD = eICU Collaborative Research Database; ICU = intensive care unit; LOS = length of stay; GBTM = group-based trajectory modeling; SHR = stress hyperglycemia ratio.

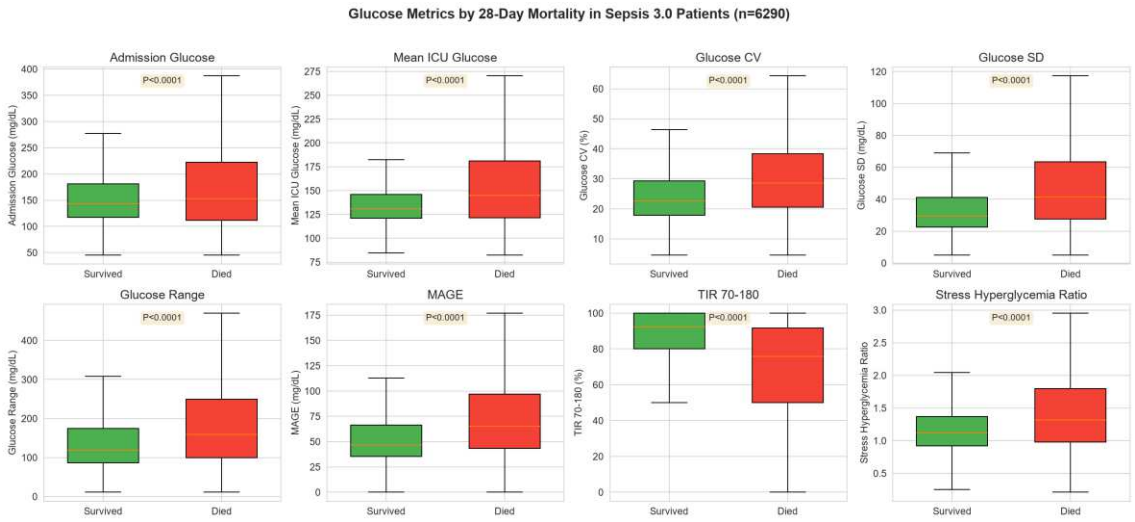


Figure 2. Box plots showing blood glucose variability measures by 28-day death status. All measures were significantly different between survivors and non-survivors ($P < 0.0001$).

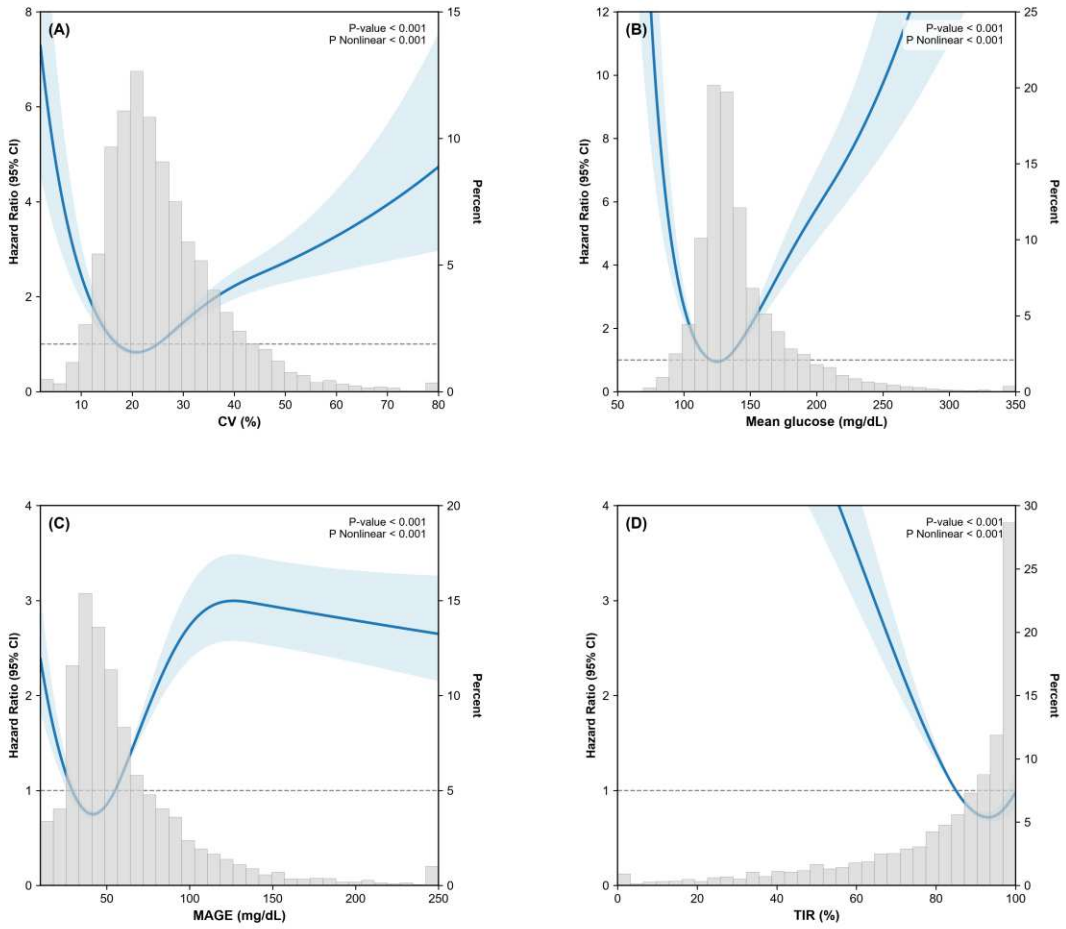


Figure 3. Restricted cubic spline curves showing the non-linear relationship between blood glucose variability measures and 28-day death risk. The shaded areas represent 95% confidence intervals. Non-linear $P < 0.0001$ for CV, mean glucose, MAGE, and TIR.

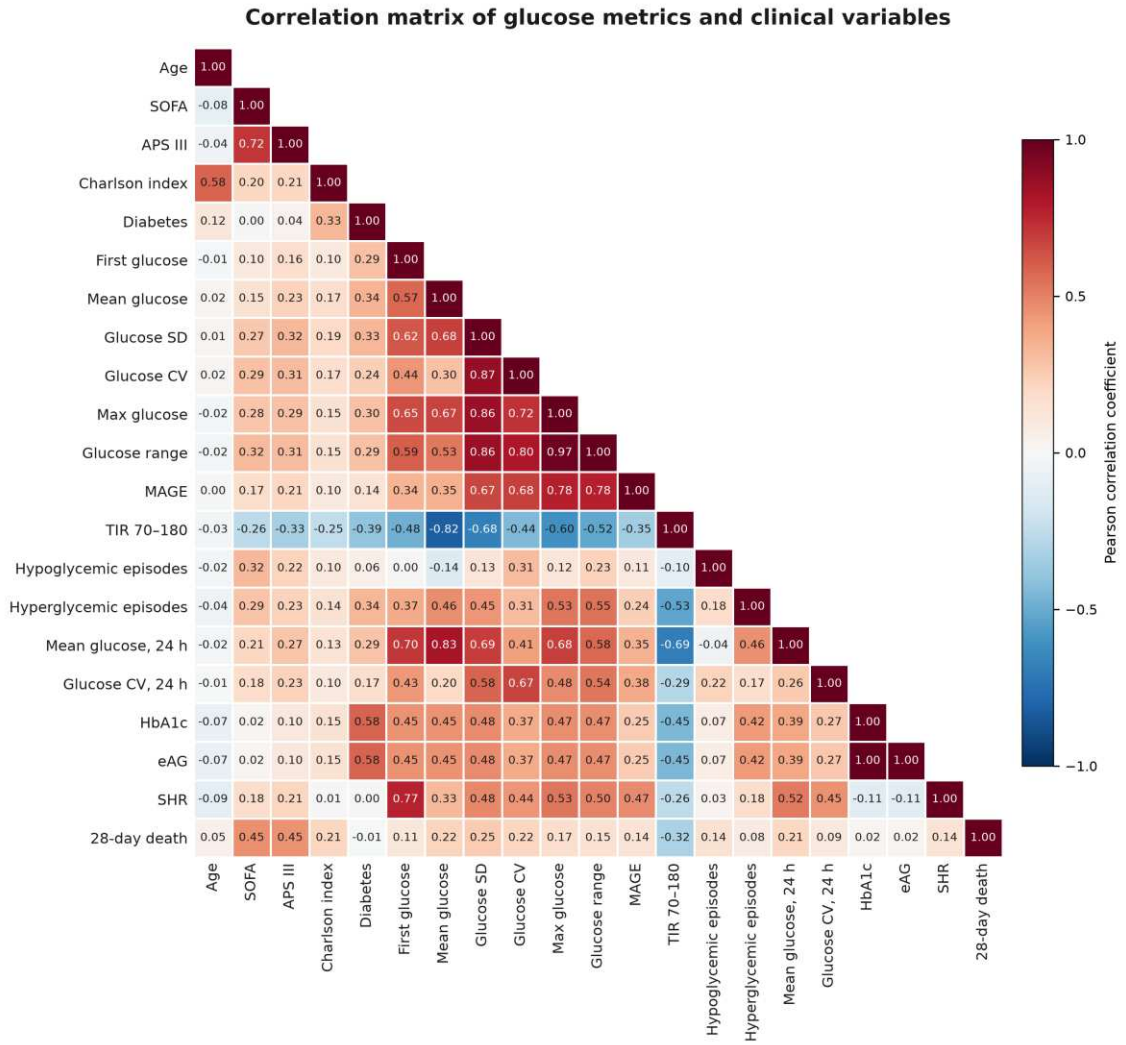


Figure 4. Correlation heatmap of blood glucose variability measures, clinical variables and 28-day death. Values are Pearson correlation coefficients; the upper triangle is left blank. CV = coefficient of variation; SD = standard deviation; MAGE = mean amplitude of glycemic excursions; TIR = time in range; HbA1c = glycated hemoglobin; eAG = estimated average glucose; SHR = stress hyperglycemia ratio.

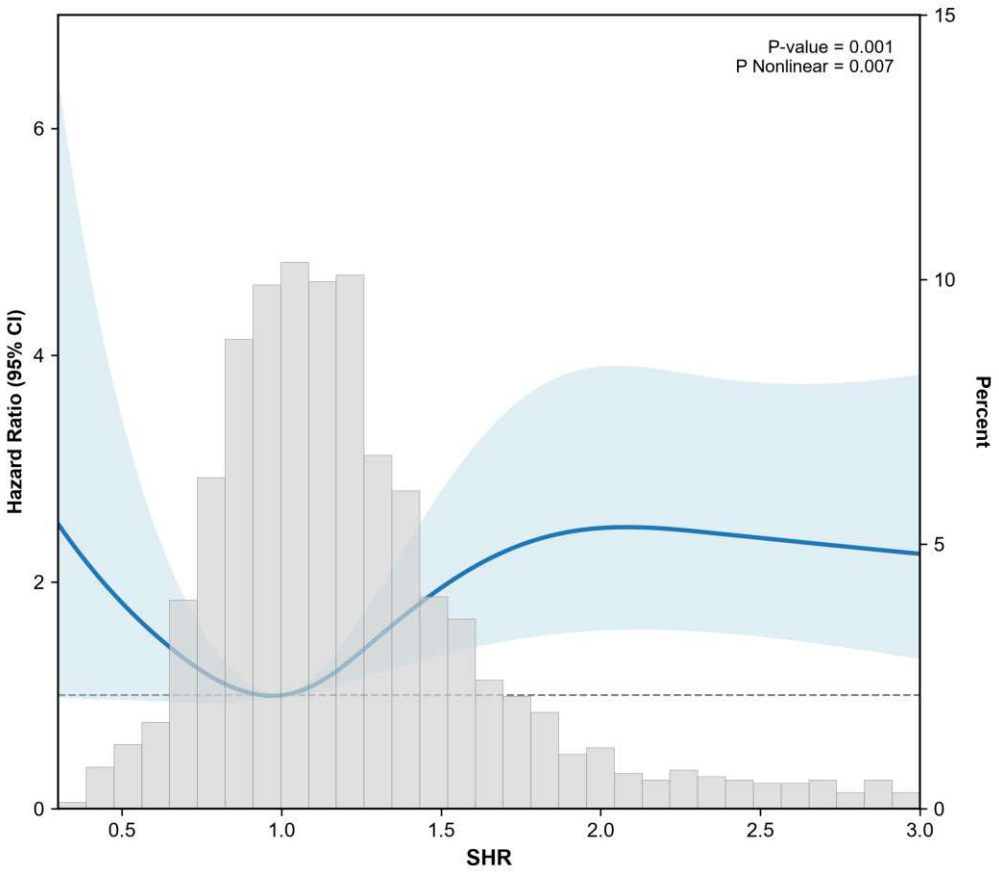


Figure 5. Restricted cubic spline curve for the association between the stress hyperglycemia ratio (SHR) and 28-day death, adjusted for age, sex, diabetes, and SOFA score. The reference line indicates HR = 1.0 at SHR = 1.0. Overall $P = 0.0014$; non-linear $P = 0.014$. Gray bars show the distribution of SHR values.

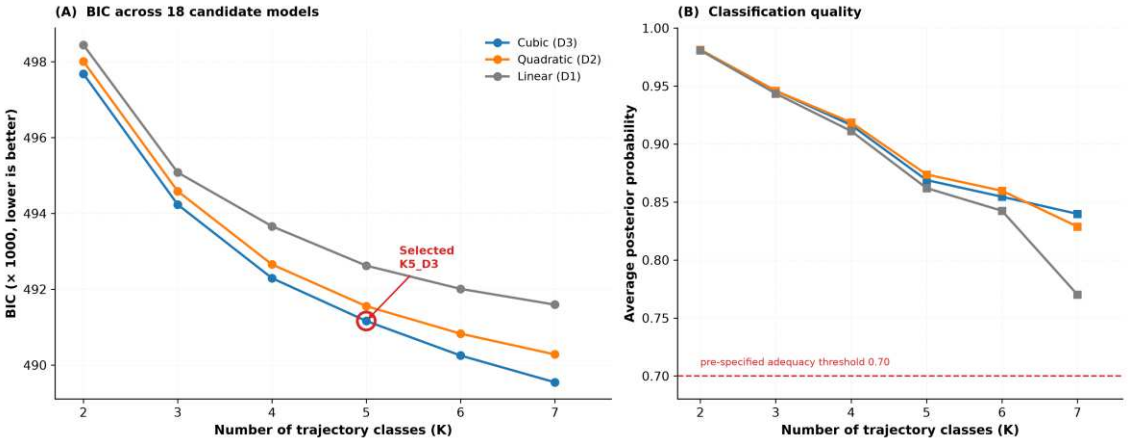


Figure 6. Model selection for group-based trajectory modeling. (A) Bayesian Information Criterion across all 18 candidate models (2–7 classes × linear/quadratic/cubic polynomials); lower is better, and the selected K5_D3 solution (five classes, cubic) is circled. BIC decreased monotonically to seven classes, but the six- and seven-class models split existing classes into subgroups with nearly identical 28-day mortality and lower classification quality. (B) Average posterior probability of assignment (AvePP); every class in the selected solution exceeded the pre-specified adequacy threshold of 0.70 (dashed line). BIC values assume a single residual variance shared across classes and are comparable with Stata *traj* and R *lcmm* but not with implementations that estimate a separate variance per class.

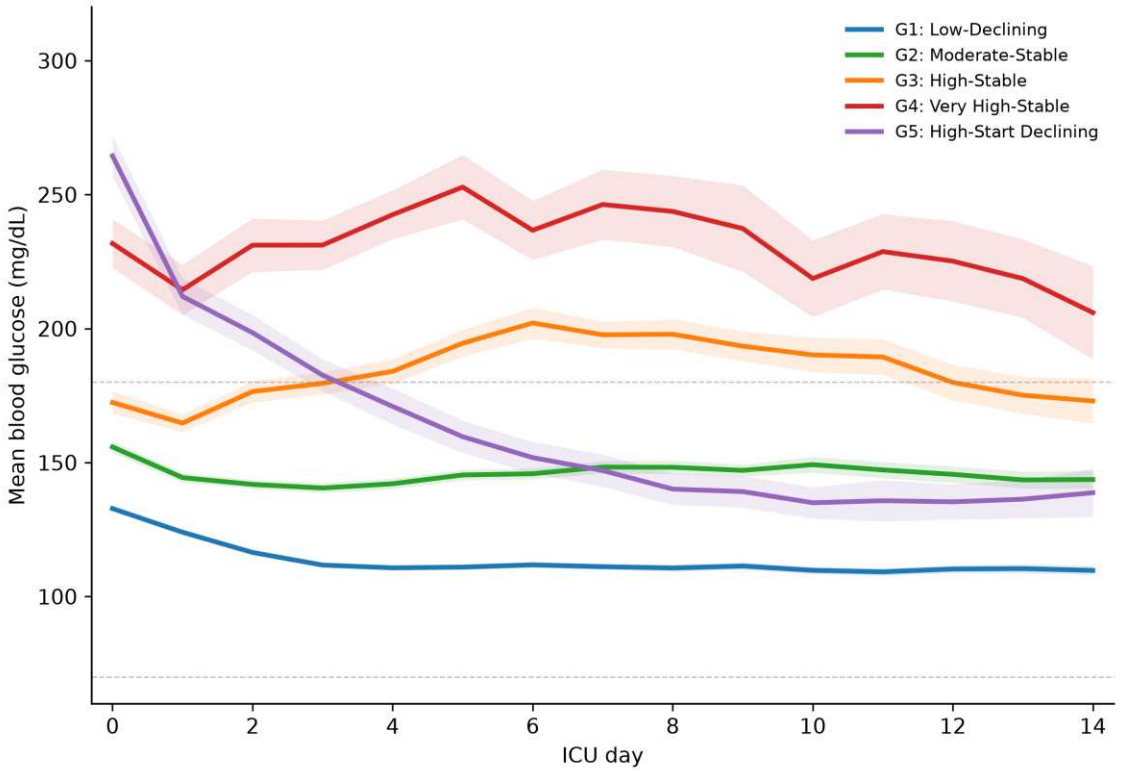


Figure 7. Blood glucose trajectory patterns identified by GBTM. The five groups show distinct blood glucose patterns over 14 ICU days. Solid lines show observed group mean daily glucose; shaded areas show 95% confidence intervals. G1: Low-Declining (blue); G2: Moderate-Stable (green); G3: High-Stable (orange); G4: Very High-Stable (red); G5: High-Start Declining (purple).

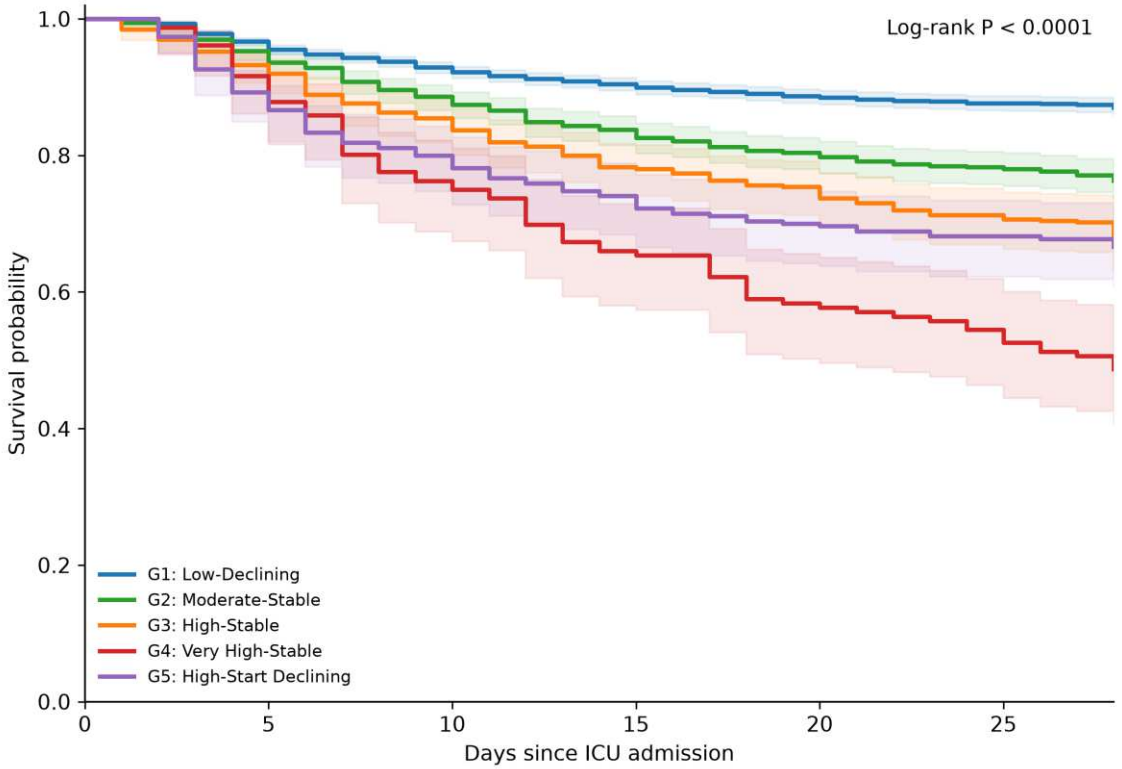


Figure 8. Kaplan-Meier survival curves for the five blood glucose trajectory groups. Log-rank test $P < 0.0001$. G4 (very high-stable) shows the lowest survival probability. Groups are color-coded as in Figure 7.

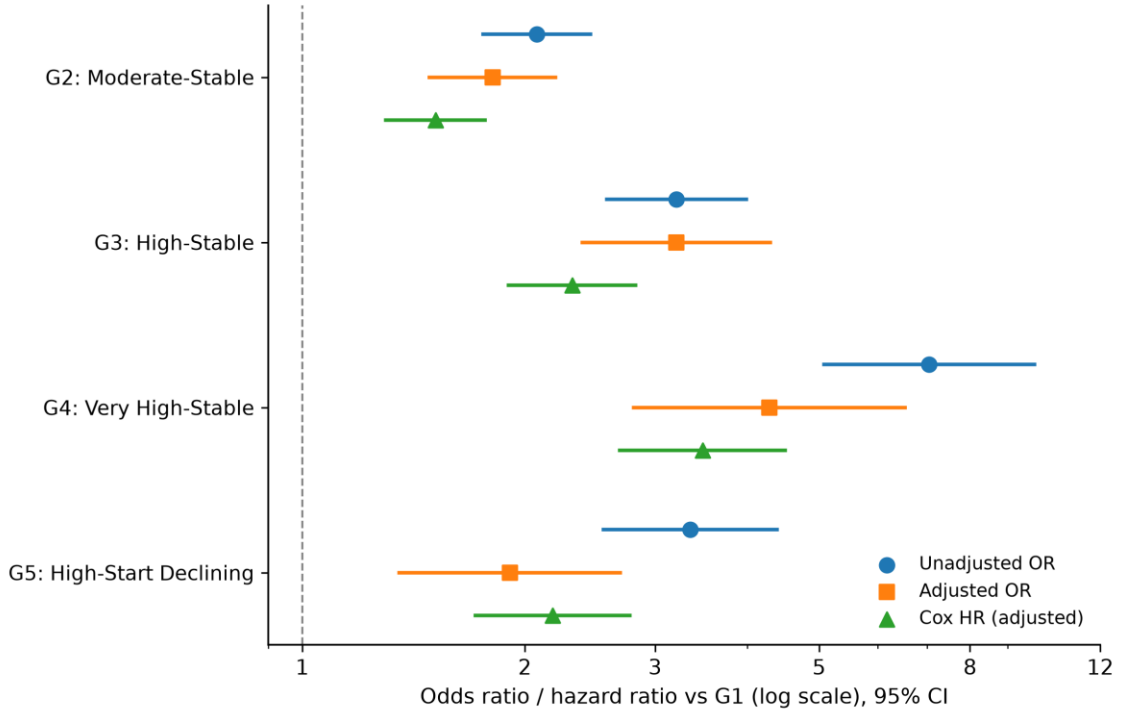


Figure 9. Forest plot showing adjusted odds ratios and hazard ratios for the association between blood glucose trajectory groups and 28-day death. Reference group: G1 (Low-Declining). Adjusted for age, sex, diabetes, and SOFA score.

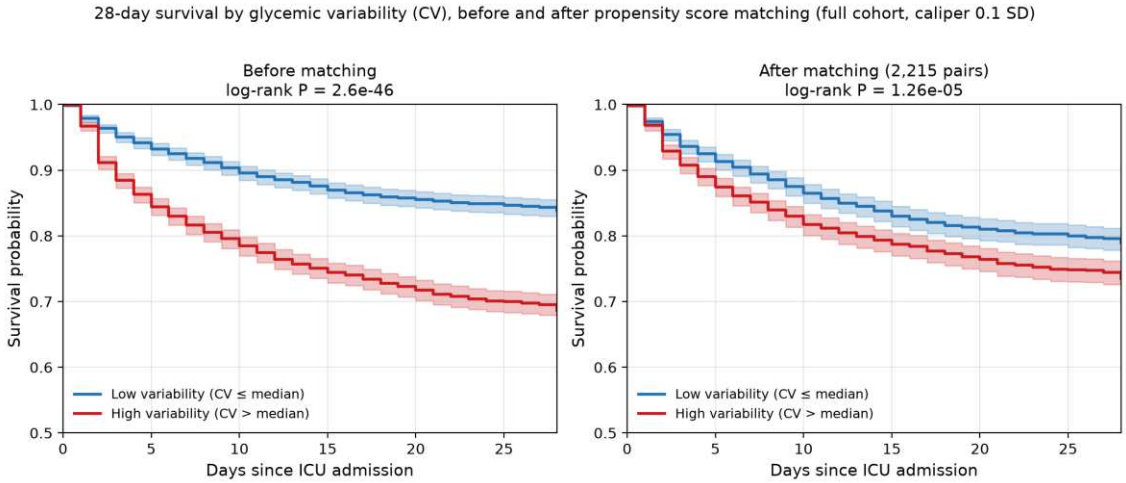


Figure 10. Propensity score matching: Kaplan-Meier curves for 28-day survival in the high- versus low-glycemic-variability (CV) groups in the full cohort of 6,290 patients, before matching (left; log-rank $P = 2.6 \times 10^{-46}$) and after 1:1 nearest-neighbour matching on the logit propensity score without replacement with a 0.1-SD caliper (right; 2,215 pairs; log-rank $P = 1.3 \times 10^{-5}$). Shaded areas are 95% confidence intervals.

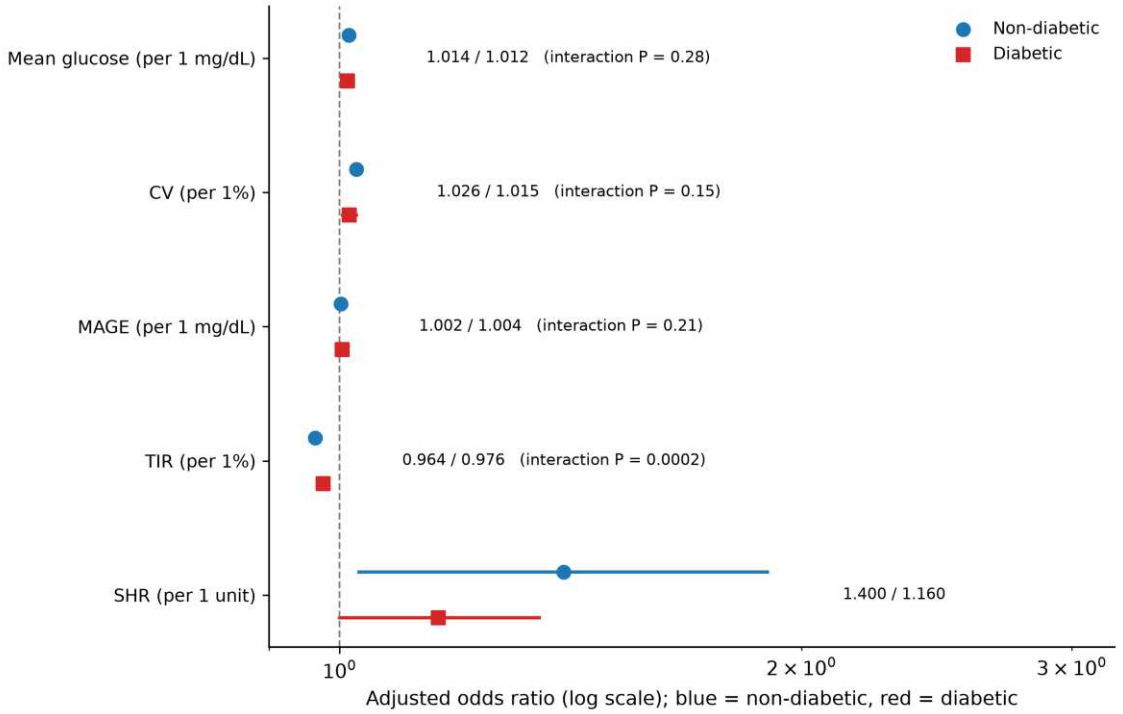


Figure 11. Diabetes-stratified forest plot showing adjusted odds ratios for blood glucose variability measures in diabetic vs non-diabetic patients.

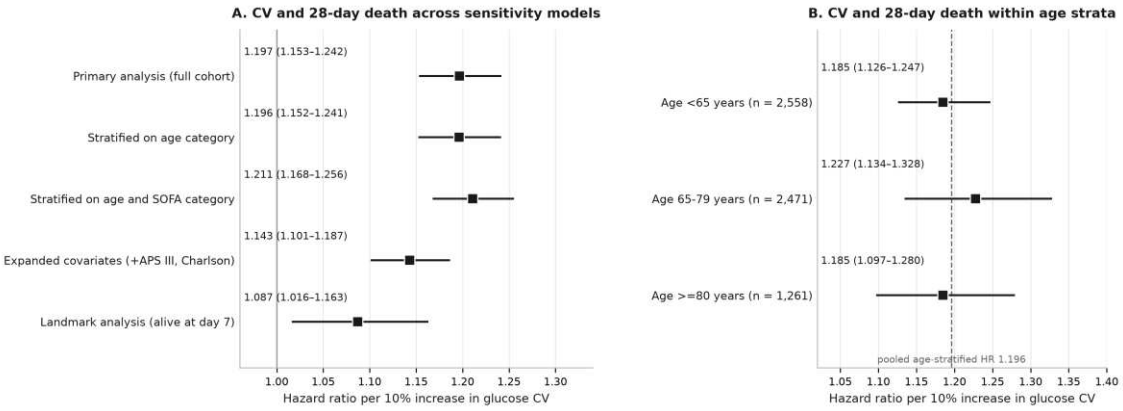


Figure 12. Sensitivity analyses of the association between glycemic variability (coefficient of variation, CV) and 28-day death. (A) Hazard ratio per 10% increase in CV across the Cox specifications used as sensitivity analyses: the primary model, models stratified on age category and on age plus SOFA category, the model expanded with APS III and the Charlson index, and the landmark analysis restricted to patients alive at day 7. (B) Hazard ratio per 10% increase in CV within each age stratum, with the pooled age-stratified estimate shown as a dashed line. All estimates are adjusted for age, sex, diabetes

and SOFA score unless the covariate is itself the stratifying variable. Squares are point estimates and horizontal lines are 95% confidence intervals.

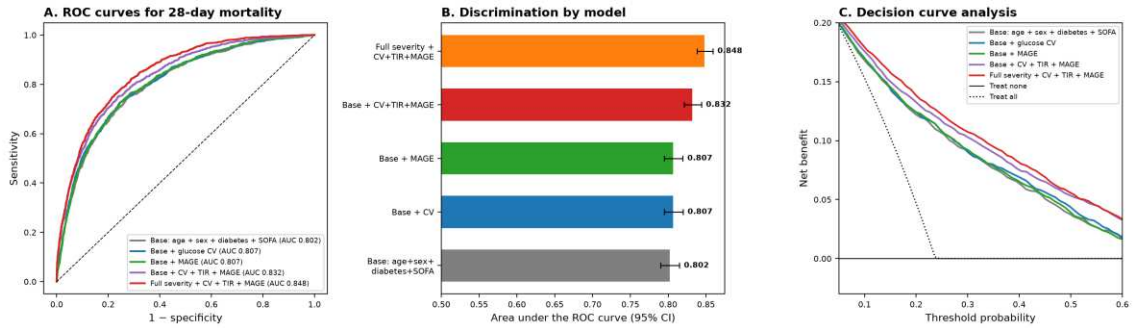


Figure 13. Incremental prognostic value of blood glucose measures for 28-day death. (A) Receiver operating characteristic curves for nested logistic regression models: the base model (age, sex, diabetes and SOFA score), the base model plus glucose coefficient of variation, plus MAGE, plus CV, time in range and MAGE, and a full severity model containing SOFA, APS III, SAPS II and the Charlson index plus the three glycemic measures. (B) Areas under the curve with 95% confidence intervals for the same models; the full severity model with glycemic measures achieved the highest discrimination (0.848). (C) Decision curve analysis: net benefit of each model across threshold probabilities from 5% to 60%, with the "treat none" and "treat all" strategies shown for reference. The model combining full severity with glycemic measures has the highest net benefit across clinically relevant thresholds. CV = coefficient of variation; MAGE = mean amplitude of glycemic excursion; AUC = area under the receiver operating characteristic curve.

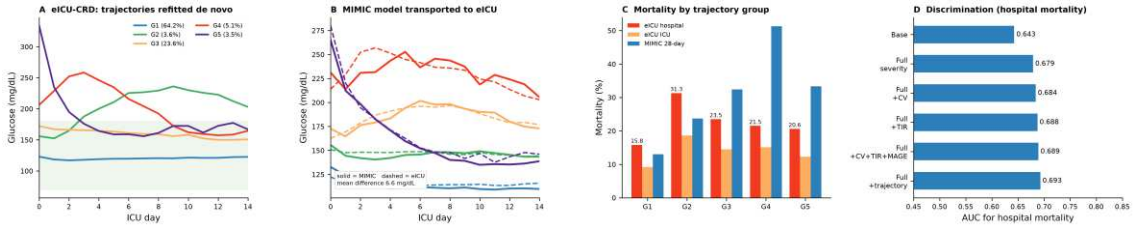


Figure 14. External validation in the eICU Collaborative Research Database (15,545 patients). (A) Five glucose trajectory groups obtained by refitting group-based trajectory modeling de novo in eICU-CRD; the shaded band is the 70-180 mg/dL target range. (B) Transport of the MIMIC-IV model: the five MIMIC-IV mean curves (solid lines) and the same curves reconstructed in eICU-CRD patients assigned with the frozen MIMIC-IV

parameters (dashed lines); the mean root-mean-square difference was 6.6 mg/dL. (C) Hospital and ICU mortality in eICU-CRD and 28-day mortality in MIMIC-IV by trajectory group; the mortality gradient was reproduced, although the highest-risk group differed between the two databases. (D) Area under the receiver operating characteristic curve for hospital mortality in eICU-CRD; time in range added more discrimination than the coefficient of variation, as in MIMIC-IV. CV = coefficient of variation; TIR = time in range; MAGE = mean amplitude of glycemic excursion; AUC = area under the receiver operating characteristic curve.

Reference

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche J-D, Coopersmith CM *et al*. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA.* 2016; 315(8):801-810. https://doi.org/10.1001/jama.2016.0287
2. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, Colombara DV, Ikuta KS, Kissoon N, Finfer S *et al*. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. *The Lancet.* 2020; 395(10219):200-211. https://doi.org/10.1016/s0140-6736(19)32989-7
3. Fleischmann-Struzek C, Mellhammar L, Rose N, Cassini A, Rudd KE, Schlattmann P, Allegranzi B, Reinhart K. Incidence and mortality of hospital- and ICU-treated sepsis: results from an updated and expanded systematic review and meta-analysis. *Intensive Care Medicine.* 2020; 46(8):1552-1562. https://doi.org/10.1007/s00134-020-06151-x
4. Fleischmann C, Scherag A, Adhikari NKJ, Hartog CS, Tsaganos T, Schlattmann P, Angus DC, Reinhart K. Assessment of global incidence and mortality of hospital-treated sepsis: current estimates and limitations. *American Journal of Respiratory and Critical Care Medicine.* 2016; 193(3):259-272. https://doi.org/10.1164/rccm.201504-0781oc
5. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, Rubenfeld G, Kahn JM, Shankar-Hari M, Singer M *et al*. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA.* 2016; 315(8):762-774. https://doi.org/10.1001/jama.2016.0288
6. Shankar-Hari M, Phillips GS, Levy ML, Seymour CW, Liu VX, Deutschman CS, Angus DC, Rubenfeld GD, Singer M. Developing a New Definition and Assessing New Clinical Criteria for Septic Shock: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA.* 2016; 315(8):775-787. https://doi.org/10.1001/jama.2016.0289
7. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, Machado FR, Mcintyre L, Ostermann M, Prescott HC *et al*. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock

2021. *Critical Care Medicine.* 2021; 49(11):e1063-e1143. https://doi.org/10.1097/ccm.0000000000005337
8. Dungan KM, Braithwaite SS, Preiser J-C. Stress hyperglycaemia. *The Lancet.* 2009; 373(9677):1798-1807. https://doi.org/10.1016/s0140-6736(09)60553-5
9. Marik PE, Bellomo R. Stress hyperglycemia: an essential survival response. *Critical Care.* 2013; 17(2):305. https://doi.org/10.1186/cc12514
10. Van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, Schetz M, Vlasselaers D, Ferdinande P, Lauwers P, Bouillon R. Intensive insulin therapy in critically ill patients. *New England Journal of Medicine.* 2001; 345(19):1359-1367. https://doi.org/10.1056/nejmoa011300
11. Umpierrez GE, Isaacs SD, Bazargan N, You X, Thaler LM, Kitabchi AE. Hyperglycemia: an independent marker of in-hospital mortality in patients with undiagnosed diabetes. *The Journal of Clinical Endocrinology & Metabolism.* 2002; 87(3):978-982. https://doi.org/10.1210/jcem.87.3.8341
12. Falciglia M, Freyberg RW, Almenoff PL, D'Alessio DA, Render ML. Hyperglycemia-related mortality in critically ill patients varies with admission diagnosis. *Critical Care Medicine.* 2009; 37(12):3001-3009. https://doi.org/10.1097/ccm.0b013e3181b083f7
13. Service FJ, Molnar GD, Rosevear JW, Ackerman E, Gatewood LC, Taylor WF. Mean amplitude of glycemic excursion, a measure of diabetic instability. *Diabetes.* 1970; 19(9):644-655. https://doi.org/10.2337/diab.19.9.644
14. Eslami S, Taherzadeh Z, Schultz MJ, Abu-Hanna A. Glucose variability measures and their effect on mortality: a systematic review. *Intensive Care Medicine.* 2011; 37(4):583-593. https://doi.org/10.1007/s00134-010-2129-5
15. Ali NA, O'Brien JM, Dungan K, Phillips G, Marsh CB, Lemeshow S, Connors AF, Preiser J-C. Glucose variability and mortality in patients with sepsis. *Critical Care Medicine.* 2008; 36(8):2316-2321. https://doi.org/10.1097/ccm.0b013e3181810378
16. Krinsley JS. Glycemic variability: a strong independent predictor of mortality in critically ill patients. *Critical Care Medicine.* 2008; 36(11):3008-3013. https://doi.org/10.1097/ccm.0b013e31818b38d2
17. Egi M, Bellomo R, Stachowski E, French Craig J, Hart G. Variability of blood glucose concentration and short-term mortality in critically ill patients. *Anesthesiology.* 2006; 105(2):244-252. https://doi.org/10.1097/00000542-200608000-00006
18. Hermanides J, Vriesendorp TM, Bosman RJ, Zandstra DF, Hoekstra JB, DeVries JH. Glucose variability is associated with intensive care unit mortality. *Critical Care Medicine.* 2010; 38(3):838-842. https://doi.org/10.1097/ccm.0b013e3181cc4be9
19. Krinsley JS, Preiser J-C. Time in blood glucose range 70 to 140 mg/dl >80% is strongly associated with increased survival in non-diabetic critically ill adults. *Critical Care.* 2015; 19(1):179. https://doi.org/10.1186/s13054-015-0908-7
20. Monnier L, Colette C. Glycemic variability: should we and can we prevent it. *Diabetes Care.* 2008; 31(Suppl 2):S150-154. https://doi.org/10.2337/dc08-s241

21. Ceriello A, Monnier L, Owens D. Glycaemic variability in diabetes: clinical and therapeutic implications. *The Lancet Diabetes & Endocrinology.* 2019; 7(3):221-230. https://doi.org/10.1016/s2213-8587(18)30136-0
22. Roberts GW, Quinn SJ, Valentine N, Alhawassi T, O'Dea H, Stranks SN, Burt MG, Doogue MP. Relative hyperglycemia, a marker of critical illness: introducing the stress hyperglycemia ratio. *The Journal of Clinical Endocrinology & Metabolism.* 2015; 100(12):4490-4497. https://doi.org/10.1210/jc.2015-2660
23. Chen L, Zeng X, Zou W, Chen M, Fan Y, Huang P. Predictive performance of stress hyperglycemia ratio for poor prognosis in critically ill patients: a systematic review and dose-response meta-analysis. *European Journal of Medical Research.* 2025; 30(1):613. https://doi.org/10.1186/s40001-025-02868-x
24. Li L, Zhou L, Peng X, Zhang Z, Zhang Z, Xiong Y, Hu Z, Yao Y. Association of stress hyperglycemia ratio and mortality in patients with sepsis: results from 13,199 patients. *Infection.* 2024; 52(5):1973-1982. https://doi.org/10.1007/s15010-024-02264-3
25. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology.* 2010; 6:109-138. https://doi.org/10.1146/annurev.clinpsy.121208.131413
26. NAGIN D: Group-Based Modeling of Development. Cambridge, MA: Harvard University Press; 2005.
27. Andrew BY, Pieper CF, Cherry AD, Pendergast JF, Privratsky JR, Mathew JP, Stafford-Smith M. Identification of trajectory-based acute kidney injury phenotypes among cardiac surgery patients. *The Annals of Thoracic Surgery.* 2022; 114(6):2235-2243. https://doi.org/10.1016/j.athoracsur.2021.11.047
28. Herle M, Micali N, Abdulkadir M, Loos R, Bryant-Waugh R, Hübel C, Bulik CM, De Stavola BL. Identifying typical trajectories in longitudinal data: modelling strategies and interpretations. *European Journal of Epidemiology.* 2020; 35(3):205-222. https://doi.org/10.1007/s10654-020-00615-6
29. Lennon H, Kelly S, Sperrin M, Buchan I, Cross AJ, Leitzmann M, Cook MB, Renehan AG. Framework to construct and interpret latent class trajectory modelling. *BMJ Open.* 2018; 8(7):e020683. https://doi.org/10.1136/bmjopen-2017-020683
30. Y Y, H W, J L, J W, L Z, C J. Relationship between blood glucose trajectory during intensive care unit stay and mortality in patients with sepsis-associated acute respiratory distress syndrome. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue.* 2025; 37(10):924-930.
31. H W, J W, J L, Y X, L Z, C J. Association between blood glucose trajectory and prognosis in septic patients in intensive care unit. 2026; 38(1):85-91.
32. Ma H, Qian X, Song X, Jiang R, Li J, Xiao F, Dou R, Guan X, Lui KY, Li S *et al*. Identifying early blood glucose trajectories in sepsis linked to distinct long-term outcomes: a K-means clustering study with external validation. *Frontiers in Immunology.* 2025; 16:1610519. https://doi.org/10.3389/fimmu.2025.1610519
33. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, Pollard TJ, Hao S, Moody B, Gow B *et al*. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data.* 2023; 10(1):1. https://doi.org/10.1038/s41597-022-01899-x

34. 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
35. Vincent J-L, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Medicine.* 1996; 22(7):707-710. https://doi.org/10.1007/bf01709751
36. Nathan DM, Kuenen J, Borg R, Zheng H, Schoenfeld D, Heine RJ. Translating the A1C assay into estimated average glucose values. *Diabetes Care.* 2008; 31(8):1473-1478. https://doi.org/10.2337/dc08-0545
37. Choi J, Hong S. The Impact of Imposing Equality Constraints on Residual Variances Across Classes in Regression Mixture Models. *Front Psychol.* 2021; 12:736132. https://doi.org/10.3389/fpsyg.2021.736132
38. McCormick EM, Byrne ML, Flournoy JC, Mills KL, Pfeifer JH. The Hitchhiker's guide to longitudinal models: A primer on model selection for repeated-measures methods. *Dev Cogn Neurosci.* 2023; 63:101281. https://doi.org/10.1016/j.dcn.2023.101281
39. Harrell FE: Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis, 2nd edn. New York: Springer; 2015.
40. 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. *Critical Care Medicine.* 2006; 34(5):1297-1310. https://doi.org/10.1097/01.ccm.0000215112.84523.f0
41. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases.* 1987; 40(5):373-383. https://doi.org/10.1016/0021-9681(87)90171-8
42. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research.* 2011; 46(3):399-424. https://doi.org/10.1080/00273171.2011.568786
43. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* 2008; 27(2):157-172; discussion 207-112. https://doi.org/10.1002/sim.2929
44. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006; 26(6):565-574. https://doi.org/10.1177/0272989x06295361
45. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet.* 2007; 370(9596):1453-1457. https://doi.org/10.1016/s0140-6736(07)61602-x
46. Roberts GW, Quinn SJ, Valentine N, Alhawassi T, O'Dea H, Stranks SN, Burt MG, Doogue MP. Relative Hyperglycemia, a Marker of Critical Illness:

Introducing the Stress Hyperglycemia Ratio. *J Clin Endocrinol Metab.* 2015; 100(12):4490-4497. https://doi.org/10.1210/jc.2015-2660
47. Wang F, Guo Y, Jiao C, Zhao S, Sui L, Mao Z, Lu R, Hou R, Zhu X. Simultaneous assessment of stress hyperglycemia ratio and glucose variability to predict all-cause mortality in sepsis patients across different glucose metabolic states: an observational cohort study with interpretable machine learning approach. *International Journal of Surgery.* 2026; 112(1):1219-1232. https://doi.org/10.1097/js9.0000000000003525
48. Monnier L, Mas E, Ginet C, Michel F, Villon L, Cristol J-P, Colette C. Activation of oxidative stress by acute glucose fluctuations compared with sustained chronic hyperglycemia in patients with type 2 diabetes. *JAMA.* 2006; 295(14):1681-1687. https://doi.org/10.1001/jama.295.14.1681
49. Ceriello A, Esposito K, Piconi L, Ihnat MA, Thorpe JE, Testa R, Boemi M, Giugliano D. Oscillating glucose is more deleterious to endothelial function and oxidative stress than mean glucose in normal and type 2 diabetic patients. *Diabetes.* 2008; 57(5):1349-1354. https://doi.org/10.2337/db08-0063
50. Turina M, Fry DE, Polk HC. Acute hyperglycemia and the innate immune system: clinical, cellular, and molecular aspects. *Critical Care Medicine.* 2005; 33(7):1624-1633. https://doi.org/10.1097/01.ccm.0000170106.61978.d8
51. Marik PE, Raghavan M. Stress-hyperglycemia, insulin and immunomodulation in sepsis. *Intensive Care Medicine.* 2004; 30(5):748-756. https://doi.org/10.1007/s00134-004-2167-y
52. Taubert D, Rosenkranz A, Berkels R, Roesen R, Sch□mig E. Acute effects of glucose and insulin on vascular endothelium. *Diabetologia.* 2004; 47(12):2059-2071. https://doi.org/10.1007/s00125-004-1586-1
53. La Sala L, Cattaneo M, De Nigris V, Pujadas G, Testa R, Bonfigli AR, Genovese S, Ceriello A. Oscillating glucose induces microRNA-185 and impairs an efficient antioxidant response in human endothelial cells. *Cardiovascular Diabetology.* 2016; 15:71. https://doi.org/10.1186/s12933-016-0390-9
54. Krinsley JS, Meyfroidt G, van den Berghe G, Egi M, Bellomo R. The impact of premorbid diabetic status on the relationship between the 3 domains of glycemic control and mortality in critically ill patients. *Current Opinion in Clinical Nutrition and Metabolic Care.* 2012; 15(2):151-160. https://doi.org/10.1097/mco.0b013e32834f0009
55. Egi M, Bellomo R, Stachowski E, French CJ, Hart GK, Taori G, Hegarty C, Bailey M. The interaction of chronic and acute glycemia with mortality in critically ill patients with diabetes. *Critical Care Medicine.* 2011; 39(1):105-111. https://doi.org/10.1097/ccm.0b013e3181feb5ea
56. Finfer S, Chittock D, Su S, Blair D, Foster D, Dhingra V, Bellomo R, Cook D, Dodek P, Henderson W *et al*. Intensive versus conventional glucose control in critically ill patients. *New England Journal of Medicine.* 2009; 360(13):1283-1297. https://doi.org/10.1056/nejmoa0810625
57. Van den Berghe G, Wilmer A, Hermans G, Meersseman W, Wouters PJ, Milants I, Van Wijngaerden E, Bobbaers H, Bouillon R. Intensive insulin therapy in the medical ICU. *New England Journal of Medicine.* 2006; 354(5):449-461. https://doi.org/10.1056/nejmoa052521

58. Finfer S, Liu B, Chittock D, Norton R, Myburgh J, McArthur C, Mitchell I, Foster D, Dhingra V, Henderson W *et al*. Hypoglycemia and risk of death in critically ill patients. *New England Journal of Medicine.* 2012; 367(12):1108-1118. https://doi.org/10.1056/nejmoa1204942
59. Egi M, Krinsley JS, Maurer P, Amin DN, Kanazawa T, Ghandi S, Morita K, Bailey M, Bellomo R. Pre-morbid glycemic control modifies the interaction between acute hypoglycemia and mortality. *Intensive Care Med.* 2016; 42(4):562-571. https://doi.org/10.1007/s00134-016-4216-8
60. Tanaka A, Yatabe T, Suhara T, Egi M. The optimal glycemic target in critically ill patients: an updated network meta-analysis. *J Intensive Care.* 2024; 12(1):14. https://doi.org/10.1186/s40560-024-00728-0
61. Egi M, Ogura H, Yatabe T, Atagi K, Inoue S, Iba T, Kakihana Y, Kawasaki T, Kushimoto S, Kuroda Y *et al*. The Japanese Clinical Practice Guidelines for Management of Sepsis and Septic Shock 2020 (J-SSCG 2020). *J Intensive Care.* 2021; 9(1):53. https://doi.org/10.1186/s40560-021-00555-7

Supplementary Files

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

- Supplementaryfile1insulinsensitivity.docx
- Supplementaryfile3sensitivityanalyses.docx
- Supplementaryfile2GBTMmodelselection.docx