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.

Longitudinal laboratory patterns and subsequent mortality in intensive care admissions with evidence of active abdominal aortic aneurysm: a MIMIC-IV landmark study

An Hongfei
Beijing Anzhen Hospital
Feng Yuan
Beijing Anzhen Hospital
Lu Jinghan
Beijing Anzhen Hospital
Geng Yihe
Beijing Anzhen Hospital
Gao Xuanzhe
Beijing Anzhen Hospital
Zhao Zezheng
Beijing Anzhen Hospital
Li Yaling
Beijing Anzhen Hospital
Wang Shuo
Beijing Anzhen Hospital
Rong Zhihua
Beijing Anzhen Hospital
Chen Zhong
Beijing Anzhen Hospital
Yang Yaoguo
yangyaoguo@ccmu.edu.cn

Beijing Anzhen Hospital

Research Article

Keywords: abdominal aortic aneurysm, intensive care unit, longitudinal laboratory patterns, landmark analysis, mortality, MIMIC-IV

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11583339/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Intensive care unit (ICU) admissions involving abdominal aortic aneurysm (AAA) are heterogeneous, and conventional summary values may obscure temporal change. We evaluated whether ordered laboratory measurements identify reproducible patterns associated with subsequent mortality in admissions with active AAA evidence.

Methods

We conducted a retrospective MIMIC-IV v3.1 landmark study. Active AAA evidence required a principal diagnosis, rupture code, repair code, or text-supported rupture or repair. Admissions alive and hospitalized at ICU hour 72 were eligible. Six laboratory markers were summarized in three consecutive 24-hour windows. Twelve posterior-sampling iterative imputations were followed by Gaussian mixture models and consensus clustering. Subsequent mortality was analyzed using logistic regression with patient-clustered robust standard errors.

Results

Of 1,343 AAA-coded ICU admissions, 526 had active AAA evidence. After excluding 16 deaths and 59 live discharges before hour 72, 451 admissions from 433 patients remained. The selected solution comprised a lower-intensity pattern (n = 289) and a multisystem derangement pattern (n = 162), with a median adjusted Rand index of 0.926. Subsequent mortality was 3.5% and 12.3%, respectively. After adjustment for age, sex, and chronic kidney disease, the odds ratio was 3.79 (95% confidence interval, 1.40-10.25; P = .009). Sensitivity estimates were directionally consistent across alternative covariate, cohort, completeness, and marker specifications. In post hoc analyses, excluding creatinine disrupted class agreement (adjusted Rand index, -0.011) and attenuated the association (odds ratio, 1.65; 95% confidence interval, 0.67–4.06); additional adjustment for early creatinine yielded an odds ratio of 2.57 (95% confidence interval, 0.87–7.58).

Conclusions

Ordered 72-hour laboratory measurements identified two reproducible patterns associated with subsequent mortality. The classification was substantially anchored by renal function and should be viewed as exploratory rather than as an AAA-specific subtype or clinical prediction tool.

Trial registration

Clinical trial number: not applicable.

Background

Abdominal aortic aneurysm (AAA) ranges from an asymptomatic anatomical finding to a catastrophic vascular emergency. Contemporary Society for Vascular Surgery and European Society for Vascular Surgery guidelines emphasize aneurysm anatomy, rupture status, symptoms, operative approach, and perioperative risk when guiding management [1, 2]. Outcomes after elective and ruptured repair have improved, but patients requiring intensive care remain heterogeneous because hemorrhage, surgery, transfusion, anticoagulation, renal hypoperfusion, contrast exposure, inflammation, and pre-existing disease may overlap within a short interval [3, 4]. Static diagnosis and procedure codes incompletely represent this evolving physiology.

Longitudinal electronic health record data may provide a complementary description of critical illness. Serial creatinine, coagulation indices, cell counts, and acid-base markers preserve information about both the level and direction of organ dysfunction. Trajectory-based subphenotyping has been applied to organ dysfunction in sepsis and to creatinine trajectories in critically ill patients with acute kidney injury [5, 6]. However, unsupervised classes can be distorted when measurements are collapsed into extrema, when missingness reflects clinical testing decisions, or when an outcome occurs before the exposure window is complete. A trajectory label is therefore credible only if within-patient ordering, missing-data uncertainty, temporal sequence, and class stability are addressed explicitly.

These challenges are pronounced in retrospective AAA studies using administrative codes. An AAA code may be secondary or incidental rather than the principal reason for admission. Diagnosis tables generally do not establish whether rupture preceded ICU admission, and procedure tables may provide only date-level timing. Consequently, laboratory changes cannot automatically be interpreted as preoperative deterioration or postoperative consequences. Adjustment for repair status may also introduce rather than remove bias when the intervention occurs within the same window used to define laboratory patterns.

We therefore redesigned the study around a restricted active-AAA evidence cohort and a 72-hour landmark. Rather than clustering maximum values and change scores, we retained ordered laboratory measurements in three consecutive ICU windows and propagated missing-data uncertainty through repeated imputation and consensus clustering. We asked whether reproducible longitudinal laboratory patterns were associated with in-hospital mortality occurring after hour 72. We prespecified that any resulting classes would be interpreted as non-specific critical-illness patterns, not AAA-specific mechanisms, causal treatment effects, or a deployable bedside algorithm.

Methods

Study design and reporting

This retrospective observational cohort study used MIMIC-IV v3.1 and was designed as an exploratory longitudinal pattern analysis with a 72-hour landmark. MIMIC-IV contains deidentified hospital and ICU data from Beth Israel Deaconess Medical Center for 2008–2022 [7, 8]. The study was structured in accordance with STROBE and RECORD reporting principles [9, 10]. The analysis protocol was finalized before interpretation of the clustering results and is supplied with the reproducible code.

Ethics and data access

MIMIC-IV is distributed through PhysioNet to credentialed investigators under a data use agreement. Source data were deidentified before release, and the original data collection received institutional approvals described by the database custodians [7, 8]. This secondary analysis used only deidentified local data and did not involve contact with patients. No attempt was made to reidentify individuals. Free-text evidence was processed locally and was not transmitted to external services.

Source AAA-coded ICU cohort

We identified hospitalizations containing ICD-9-CM codes 441.3 or 441.4 or ICD-10-CM codes I71.3, I71.30-I71.33, I71.4, or I71.40-I71.43 after removal of punctuation. Hospitalizations were linked to ICU stays, and the first ICU stay within each eligible hospitalization was retained. Adults were eligible; all linked records had adult age values. The unit of analysis was the hospitalization-specific first ICU admission. Repeated eligible hospitalizations from the same patient were retained and addressed using patient-clustered inference and a first-admission sensitivity analysis.

Rupture was approximated using ICD-9-CM code 441.3 or the corresponding ruptured ICD-10-CM codes. AAA repair was approximated from ICD procedure codes, including ICD-9-CM 39.71 for endovascular repair and 38.44 for open abdominal aortic replacement, together with prespecified ICD-10-PCS patterns for abdominal aortic restriction, replacement, or supplement procedures. These variables were not treated as causal treatment exposures.

Active AAA evidence

The primary analysis required at least one of five indicators that AAA was clinically active during the hospitalization: (1) an AAA diagnosis at sequence position 1, (2) a ruptured AAA code, (3) an AAA repair code, (4) text-supported rupture, or (5) text-supported AAA repair. Sequence position 1 was interpreted as the principal coded diagnosis. Text support was derived from locally available MIMIC-IV-Note discharge summaries and radiology reports [11]. Prespecified non-negated expressions identified rupture, endovascular aneurysm repair, or open repair. Text support was used as evidence enrichment rather than manual clinical adjudication. Admissions lacking notes could still qualify through principal diagnosis, rupture, or repair codes.

Landmark cohort

ICU admission was the time origin. The exposure window extended from ICU hour 0 through hour 72. To ensure that the outcome followed completion of this window, the primary landmark cohort included only active-evidence admissions in which the patient was alive and remained hospitalized at hour 72. Patients who died before the landmark and patients discharged alive before the landmark were reported separately and excluded from the primary association analysis. The primary outcome was in-hospital death after hour 72, based on the hospital expiration indicator.

Laboratory extraction and longitudinal representation

Laboratory events were obtained from the MIMIC-IV hospital laboratory table using prespecified item identifiers. The primary markers were creatinine, international normalized ratio (INR), platelet count, white blood cell count, hemoglobin, and bicarbonate. These markers represented renal function, coagulation, hematologic state, inflammation or physiological stress, anemia or bleeding-related physiology, and acid-base state. Lactate was retained for descriptive analyses but excluded from primary clustering because its availability fell markedly in later windows and was likely driven by clinical testing decisions.

For each admission and marker, the median observed value was calculated within ICU hours 0–24, 24–48, and 48–72. These ordered window values formed an 18-cell patient-by-marker-by-time matrix. Pre-ICU measurements were not included because their source could include the emergency department, operating room, ward, or interhospital transfer and therefore did not share a uniform ICU anchor. The 0.5th and 99.5th percentiles of each marker were used to limit extreme values. Creatinine, INR, and white blood cell values were transformed using $\log(1 + x)$; platelet count, hemoglobin, and bicarbonate remained on their original scales. Robust scaling was applied before imputation and clustering.

Missing data and repeated imputation

Missingness was evaluated for every marker and time window. Outcomes were not used during imputation or clustering. The longitudinal matrix was augmented with age, sex, aggregate comorbidity count, principal AAA diagnosis, rupture code, and repair code solely as imputation predictors. Twelve iterative multivariable imputations were generated with Bayesian ridge regression, posterior sampling, 25 iterations, and distinct fixed random seeds. This procedure propagated imputation uncertainty across class discovery rather than treating one median-filled matrix as complete. Because clinically driven testing may be missing not at random, repeated imputation was treated as an uncertainty analysis rather than proof that missingness bias had been eliminated [12, 13].

Consensus clustering and class selection

Within each imputed dataset, the 18 longitudinal cells were standardized and fitted using diagonal-covariance Gaussian mixture models with 2–5 candidate classes, 30 initializations, a maximum of 1,000 iterations, and covariance regularization. We recorded the Bayesian information criterion (BIC), Akaike information criterion, silhouette coefficient, convergence, and minimum class size for every imputation and candidate class count.

The primary selection rule required every class to contain at least 10% of the cohort and prioritized the highest median silhouette coefficient; BIC was used as a secondary diagnostic. This rule was chosen because information criteria may continue to improve as clinically weak classes are subdivided. For the selected class count, an admission-level co-assignment matrix was calculated across the 12 imputations. Average-linkage hierarchical clustering of one minus the co-assignment probability produced the final consensus classes [14]. Agreement between every imputation-specific assignment and the final assignment was quantified using the adjusted Rand index. Mortality was not used to create, order, select, or name classes. Classes were ordered using a prespecified laboratory severity score that assigned higher severity to higher creatinine, INR, and white blood cell values and to lower platelet, hemoglobin, and bicarbonate values.

Longitudinal descriptions

For display, trajectory values were calculated from observed, non-imputed window medians. For every class, marker, and time window, we reported the median, interquartile range, number observed, and a 95% nonparametric bootstrap confidence interval based on 1,000 samples. Population-averaged generalized estimating equations (GEE) with an exchangeable within-admission correlation structure described class differences in standardized marker level and linear change across the three windows [15, 16]. Because the same marker trajectories contributed to class assignment, GEE interaction tests were interpreted descriptively rather than as independent validation.

Baseline variables

Baseline variables included age, sex, and diagnosis-code indicators for hypertension, diabetes mellitus, chronic kidney disease, heart failure, coronary artery disease, chronic obstructive pulmonary disease, and peripheral artery disease. Age was capped according to MIMIC-IV deidentification conventions. Chronic kidney disease was selected a priori for the primary adjusted outcome model because it is a specific baseline determinant of creatinine trajectories.

Outcome analysis

Subsequent mortality was summarized as events, proportions, and Wilson 95% confidence intervals. A logistic generalized linear model estimated the association between longitudinal pattern and subsequent in-hospital death after adjustment for age, sex, and pre-existing chronic kidney disease. Robust standard errors were clustered by patient to account for repeated hospitalizations. Results are reported as odds ratios (ORs), 95% confidence intervals (CIs), and two-sided P values.

Prespecified sensitivity analyses (1) additionally adjusted for rupture and repair codes; (2) replaced chronic kidney disease with aggregate comorbidity count; (3) restricted the cohort to admissions with a principal AAA diagnosis; (4) retained only the first eligible hospitalization per patient; (5) required at least 15 of 18 longitudinal laboratory cells to be observed; and (6) repeated two-class imputation and consensus clustering after excluding INR. The final analysis evaluated both agreement with the primary assignment and the adjusted mortality association. Two post hoc analyses examined dependence on the renal signal. The first repeated imputation and consensus clustering after excluding creatinine. The second added the observed median creatinine during hours 0–24, transformed as log(1 + x), to the primary outcome model; admissions without an observed early creatinine value were excluded from this model.

Software and reproducibility

Analyses were performed in Python 3.12.7 using pandas 2.2.2, NumPy, SciPy 1.13.1, scikit-learn 1.5.1, and statsmodels 0.14.2 [17, 18]. Figures were produced with matplotlib 3.9.2. Random seeds, transformations, candidate models, imputation-specific diagnostics, assignments, source-data tables, and figure-generation code are included in the analysis package. No multiplicity adjustment was applied because one primary outcome association was prespecified; all longitudinal marker tests and sensitivity analyses were exploratory.

Results

Cohort construction and admission context

The source cohort contained 1,343 AAA-coded ICU admissions. Only 395 (29.4%) had an AAA diagnosis at sequence position 1, confirming that a broad code-defined cohort included many admissions in which AAA might not have been the principal condition. The active-evidence definition retained 526 admissions. Sixteen patients died before ICU hour 72 and 59 were discharged alive before hour 72. The primary landmark cohort therefore comprised 451 admissions from 433 patients (Fig. 1a).

Within the landmark cohort, 341 admissions (75.6%) had a principal AAA diagnosis, 89 (19.7%) had a ruptured AAA code, 327 (72.5%) had an AAA repair code, 112 (24.8%) had text-supported rupture, and 272 (60.3%) had text-supported repair; these categories overlapped (Fig. 1b). Emergency and same-day surgical admissions accounted for 166 (36.8%) and 162 (35.9%) admissions, respectively. A further 89 admissions (19.7%) originated as transfers from another hospital. Thus, the primary cohort was substantially more vascular-specific than the original code-defined cohort but remained clinically heterogeneous.

Median age was 74 years (interquartile range [IQR], 67–80), and 150 patients (33.3%) were female. Hypertension was recorded in 373 admissions (82.7%), chronic kidney disease in 117 (25.9%), coronary artery disease in 195 (43.2%), heart failure in 100 (22.2%), chronic obstructive pulmonary disease in 132 (29.3%), and peripheral artery disease in 115 (25.5%) (Table 1).

Table 1
Baseline and admission characteristics by longitudinal laboratory pattern

Characteristic	Overall (n = 451)	Lower-intensity pattern (n = 289)	Multisystem derangement pattern (n = 162)
Age, years	74.0 (67.0–80.0)	73.0 (66.0–80.0)	75.0 (69.0-80.8)
Female sex	150/451 (33.3%)	103/289 (35.6%)	47/162 (29.0%)
Principal AAA diagnosis	341/451 (75.6%)	221/289 (76.5%)	120/162 (74.1%)
Ruptured AAA code	89/451 (19.7%)	40/289 (13.8%)	49/162 (30.2%)
Any AAA repair code	327/451 (72.5%)	212/289 (73.4%)	115/162 (71.0%)
Hypertension	373/451 (82.7%)	238/289 (82.4%)	135/162 (83.3%)
Diabetes mellitus	82/451 (18.2%)	44/289 (15.2%)	38/162 (23.5%)
Chronic kidney disease	117/451 (25.9%)	29/289 (10.0%)	88/162 (54.3%)
Heart failure	100/451 (22.2%)	48/289 (16.6%)	52/162 (32.1%)
Coronary artery disease	195/451 (43.2%)	111/289 (38.4%)	84/162 (51.9%)
Chronic obstructive pulmonary disease	132/451 (29.3%)	81/289 (28.0%)	51/162 (31.5%)
Peripheral artery disease	115/451 (25.5%)	63/289 (21.8%)	52/162 (32.1%)
Aggregate comorbidity count	2.0 (1.0–3.0)	2.0 (1.0–3.0)	3.0 (2.0–4.0)

Values are median (interquartile range) or n/N (%). AAA, abdominal aortic aneurysm.

Longitudinal data availability

Creatinine, platelet count, white blood cell count, hemoglobin, and bicarbonate were observed in at least 99% of admissions during hours 0–24 and in 95%-97% during hours 48–72 (Fig. 1c). INR availability declined from 89% to 57%. Lactate availability declined from 76% during hours 0–24 to 42% during hours 24–48 and 27% during hours 48–72. This pattern supported excluding lactate from primary clustering and retaining it only for descriptive interpretation.

Selection and stability of longitudinal patterns

All candidate models converged. Median BIC decreased from 21,962 for two classes to 20,998 for five classes, but median silhouette coefficient simultaneously declined from 0.153 to 0.055. The five-class solution also produced a minimum class below the prespecified 10% threshold. The two-class solution met the minimum-size rule and had the highest median silhouette coefficient and was therefore selected (Fig. 2h). The modest absolute silhouette value indicated that heterogeneity was partly continuous rather than divided into sharply separated biological subtypes.

Consensus clustering assigned 289 admissions (64.1%) to the lower-intensity pattern and 162 (35.9%) to the multisystem derangement pattern. Co-assignment was strong within the two consensus blocks (Fig. 2g). The median adjusted Rand index comparing the 12 imputation-specific solutions with the consensus assignment was 0.926 (range, 0.742–0.973), indicating that class membership was generally stable to imputation uncertainty (Fig. 2i).

Observed longitudinal profiles

The lower-intensity pattern showed a median creatinine of 0.9 mg/dL in all three windows, bicarbonate rising from 23.5 to 25.0 mEq/L, platelet counts decreasing from 152 to 134 and then 137 x10^3/μL, and hemoglobin decreasing from 10.8 to 10.0 g/dL. Median INR remained 1.2. The multisystem derangement pattern showed higher creatinine (1.65, 2.00, and 2.00 mg/dL), lower bicarbonate (21, 22, and 22 mEq/L), lower hemoglobin (9.9, 9.45, and 9.5 g/dL), lower platelet counts (135, 121.5, and 119 x10^3/μL), and higher early INR (1.3, 1.3, and 1.2) (Fig. 2a-f; Table 2). Lactate, which was not used for class discovery, was also higher in the multisystem derangement pattern among admissions with observed values.

Table 2
Observed longitudinal laboratory values by pattern and ICU window

Marker / pattern	0–24 h	24–48 h	48–72 h
Creatinine, mg/dL — Lower-intensity pattern	0.9 (0.75–1.1); n = 288	0.9 (0.75–1.1); n = 281	0.9 (0.7–1.1); n = 277
Creatinine, mg/dL — Multisystem derangement pattern	1.65 (1.3–2.15); n = 161	2 (1.55–2.65); n = 161	2 (1.6–2.9); n = 159
INR — Lower-intensity pattern	1.2 (1.1–1.3); n = 251	1.2 (1.1–1.3); n = 182	1.2 (1.1–1.3); n = 145
INR — Multisystem derangement pattern	1.3 (1.2–1.5); n = 150	1.3 (1.1–1.6); n = 128	1.2 (1.1–1.5); n = 114
Platelet count, ×10³/μL — Lower-intensity pattern	152 (120-193.5); n = 289	134 (103.5–177); n = 279	137 (105-174.25); n = 272
Platelet count, ×10³/μL — Multisystem derangement pattern	135 (102–181); n = 161	121.5 (92.75-163.62); n = 160	119 (94.75-165.25); n = 156
White blood cell count, ×10³/μL — Lower-intensity pattern	10.3 (8.3-12.85); n = 289	10.5 (8.4–12.6); n = 279	9.8 (8-12.15); n = 271
White blood cell count, ×10³/μL — Multisystem derangement pattern	10.6 (7.9–14.3); n = 161	10.8 (8.53–14.62); n = 160	10 (7.97–13.86); n = 156
Hemoglobin, g/dL — Lower-intensity pattern	10.8 (9.8–11.9); n = 289	10.2 (9.12–11.3); n = 278	10 (9.1–11); n = 271
Hemoglobin, g/dL — Multisystem derangement pattern	9.9 (9.1–10.9); n = 161	9.45 (8.79–10.51); n = 160	9.5 (8.6–10.2); n = 156
Bicarbonate, mEq/L — Lower-intensity pattern	23.5 (21.5–25); n = 288	25 (23–27); n = 281	25 (23–27); n = 275
Bicarbonate, mEq/L — Multisystem derangement pattern	21 (19–23); n = 161	22 (20–24); n = 161	22 (20.5–24.5); n = 157
Lactate, mmol/L (descriptive only) — Lower-intensity pattern	1.7 (1.15–2.4); n = 207	1.2 (1-1.65); n = 92	1.05 (0.9–1.38); n = 58
Lactate, mmol/L (descriptive only) — Multisystem derangement pattern	2.1 (1.27–3.25); n = 135	1.55 (1.14–2.21); n = 96	1.4 (1-2.02); n = 63

Values are median (interquartile range); n = number of admissions with an observed window median. Lactate was not used for clustering.

In the descriptive longitudinal GEE models, the multisystem derangement pattern had a greater standardized creatinine increase per 24-hour window than the lower-intensity pattern (interaction estimate, 0.238 standard deviations; 95% CI, 0.180–0.297; P < .001). Bicarbonate recovery was slower (interaction estimate, -0.126 standard deviations; 95% CI, -0.214 to -0.039; P = .005). Time-by-pattern interactions for INR, platelet count, white blood cell count, and hemoglobin were not statistically distinct,

indicating that several class differences primarily reflected persistent levels rather than divergent linear slopes.

Patient characteristics by longitudinal pattern

Age and sex distributions were similar between patterns, but baseline comorbidity differed. Chronic kidney disease was recorded in 29 of 289 lower-intensity admissions (10.0%) and 88 of 162 multisystem derangement admissions (54.3%). Heart failure occurred in 16.6% and 32.1%, and coronary artery disease in 38.4% and 51.9%, respectively. Ruptured AAA codes occurred in 13.8% of the lower-intensity pattern and 30.2% of the multisystem derangement pattern. Any AAA repair code occurred in 73.4% and 71.0%, respectively (Table 1). These imbalances motivated specific adjustment for chronic kidney disease and sensitivity adjustment for rupture and repair.

Subsequent mortality

Thirty patients died in hospital after the 72-hour landmark. Subsequent mortality was 3.5% (10/289; 95% CI, 1.9%-6.2%) in the lower-intensity pattern and 12.3% (20/162; 95% CI, 8.1%-18.3%) in the multisystem derangement pattern (Fig. 3a). Median ICU length of stay was 2.27 and 4.11 days, and median hospital length of stay was 7.51 and 10.69 days, respectively.

After adjustment for age, sex, and pre-existing chronic kidney disease, the multisystem derangement pattern was associated with higher subsequent in-hospital mortality (OR, 3.79; 95% CI, 1.40-10.25; P = .009) (Fig. 3b; Table 3). Chronic kidney disease itself was not independently associated with the outcome in this model (OR, 1.12; 95% CI, 0.43–2.90), although the limited number of events produced imprecise estimates.

Table 3
Primary adjusted association with subsequent in-hospital mortality

Variable	Adjusted OR	95% CI	P value
Multisystem derangement vs lower-intensity pattern	3.79	1.40-10.25	.009
Age, per year	1.04	0.99–1.09	.152
Female sex	1.69	0.78–3.67	.185
Chronic kidney disease	1.12	0.43–2.90	.822

Logistic generalized linear model with patient-clustered robust standard errors; N = 451 admissions, 30 deaths. OR, odds ratio; CI, confidence interval.

Sensitivity analyses

Additional adjustment for rupture and repair codes attenuated the pattern association but preserved its direction and statistical compatibility with increased mortality (OR, 2.93; 95% CI, 1.10–7.82; P = .031). An alternative model adjusting for age, sex, and aggregate comorbidity count produced an OR of 3.19 (95%

CI, 1.22–8.36; P = .018). Restriction to 341 principal-AAA admissions produced an OR of 4.83 (95% CI, 1.26–18.49; P = .021). Restriction to the first eligible admission per patient produced an OR of 4.81 (95% CI, 1.51–15.32; P = .008). Among 417 admissions with at least 15 of 18 longitudinal cells observed, the OR was 3.52 (95% CI, 1.32–9.35; P = .012) (Fig. 3c; Table 4).

Table 4
Sensitivity analyses for subsequent in-hospital mortality

Analysis	N	Deaths	Adjusted OR (95% CI)	P value
Primary model	451	30	3.79 (1.40-10.25)	.009
Additional rupture and repair covariates	451	30	2.93 (1.10–7.82)	.031
Age, sex, and aggregate comorbidity burden	451	30	3.19 (1.22–8.36)	.018
Principal AAA diagnosis only	341	21	4.83 (1.26–18.49)	.021
First eligible admission per patient	433	26	4.81 (1.51–15.32)	.008
At least 15 of 18 longitudinal cells observed	417	30	3.52 (1.32–9.35)	.012
Two-class consensus clustering without INR	451	30	3.45 (1.32–9.05)	.012
Post hoc: two-class consensus clustering without creatinine	451	30	1.65 (0.67–4.06)	.280
Post hoc: primary pattern with additional adjustment for 0–24 h creatinine	449	30	2.57 (0.87–7.58)	.087

All estimates compare the multisystem derangement pattern with the lower-intensity pattern and adjust for age, sex, and chronic kidney disease unless otherwise stated. These are association estimates, not causal effects.

When INR was excluded from clustering, the two-class solution was highly stable across imputations (median adjusted Rand index, 0.982) and showed moderate agreement with the primary classification (adjusted Rand index, 0.653). The adjusted mortality association remained similar (OR, 3.45; 95% CI, 1.32–9.05; P = .012). These analyses reduced, but did not eliminate, concern that INR missingness or an isolated coagulation signal determined the primary result.

The post hoc renal-signal analyses produced a different result. After creatinine was excluded, the two-class solution remained moderately stable across imputations (median adjusted Rand index, 0.804) but had no agreement with the primary classification beyond chance (adjusted Rand index, -0.011). Its adjusted mortality association was attenuated and imprecise (OR, 1.65; 95% CI, 0.67–4.06; P = .280). Additional adjustment of the primary classification for observed 0-24-hour creatinine included 449 admissions and produced an OR of 2.57 (95% CI, 0.87–7.58; P = .087). These post hoc results indicate

that creatinine was a major determinant of both class membership and the observed outcome association.

Discussion

This landmark analysis identified two reproducible longitudinal laboratory patterns among ICU admissions with evidence that AAA was clinically active. The lower-intensity pattern maintained near-normal creatinine and showed improving bicarbonate. The multisystem derangement pattern combined renal dysfunction with lower bicarbonate, hemoglobin, and platelet counts and modest coagulation abnormality. Because mortality was assessed only after the three measurement windows were complete, the primary association had a coherent temporal order. Post hoc analyses nevertheless showed that renal function was central to both the classification and its association with mortality.

The main contribution is methodological rather than a new bedside score. Approaches that compress a 72-hour window into extrema and change scores do not preserve the ordered patient-level structure. The present design retained three consecutive windows, propagated uncertainty across 12 imputations, and based membership on consensus rather than a single completed dataset. Similar trajectory approaches have described heterogeneity in sepsis-related organ dysfunction and acute kidney injury [5, 6]. MIMIC-based studies in aortic aneurysm have instead compared established severity scores or evaluated a single hematologic marker [19, 20]. Our analysis therefore adds an explicitly time-ordered, multivariable description while showing that the resulting structure remains strongly renal-function dependent.

The study also clarifies the boundary between vascular context and general critical illness. Most landmark admissions had a principal AAA diagnosis or repair evidence, and the association strengthened when analysis was restricted to principal AAA diagnoses. Nevertheless, the laboratory pattern is not demonstrably AAA-specific. Higher creatinine and lower bicarbonate may reflect chronic kidney disease, shock, contrast exposure, perioperative stress, infection, or other organ dysfunction. Lower platelet counts and hemoglobin may reflect bleeding, hemodilution, transfusion, consumption, or pre-existing disease. INR may be influenced by anticoagulants, liver dysfunction, plasma administration, or critical illness. These alternative explanations cannot be separated reliably without exact procedure times, medication administration, transfusion data, and richer clinical adjudication.

The marked imbalance in chronic kidney disease prevalence required particular attention. Adjustment for coded chronic kidney disease did not remove the primary association, but a diagnosis code cannot capture baseline renal function or acute renal physiology. Removing creatinine produced a substantially different classification, and adjustment for early creatinine attenuated the association. The primary pattern should therefore be interpreted as a renal-function-anchored composite of physiological derangement, not as evidence of a renal-independent multisystem subtype. Residual severity confounding also remains plausible because the analysis did not include Glasgow Aneurysm Score, Vascular POSSUM, SOFA, APACHE, SAPS, vasopressor dose, mechanical ventilation, transfusion volume, pH, base deficit, operative urgency, or aneurysm anatomy.

The landmark design improves temporal ordering but changes the target population. Sixteen patients who died and 59 who were discharged before hour 72 were excluded. The reported OR therefore applies to patients who remained hospitalized and alive after the complete measurement window, not to all AAA admissions and not to admission-time mortality prediction. Early discharge may also create informative missingness before the landmark, which is why patients discharged before hour 72 were excluded rather than having late values imputed. The resulting patterns cannot be calculated prospectively at ICU admission because all three windows are required.

Missingness remains another important limitation. Lactate measurement was strongly time-dependent and clinically driven, so lactate was removed from primary clustering. INR remained less complete than the other markers, although the no-INR analysis yielded a similar mortality association. Repeated imputation with posterior sampling propagated uncertainty under an imputation model, but it could not identify or correct all missing-not-at-random mechanisms. The modest silhouette coefficient further indicates that the two classes simplify a continuum of physiological derangement; they should not be interpreted as discrete biological entities.

Other limitations include the single-center design, code- and rule-based active AAA definition, incomplete note availability, date-level rather than time-stamped operative coding, and only 30 subsequent deaths. Patient-clustered standard errors addressed repeated admissions, but confidence intervals remained wide. The renal-signal analyses were post hoc and should be interpreted as diagnostic checks rather than confirmatory tests. Automated text patterns were not equivalent to manual chart adjudication. Laboratory values could mix preoperative and postoperative physiology because the ICU anchor did not establish exact procedure sequence. Finally, the analysis lacks external validation and prospective evaluation.

The appropriate next step is external validation in a vascular registry linked to ICU data. Such a cohort could adjudicate aneurysm anatomy, symptoms, rupture at presentation, operative timing, transfusion, anticoagulation, organ support, and baseline renal function. Validation should test whether the two-pattern solution reproduces without refitting and whether it adds information beyond renal measures and established vascular and ICU severity scores. A clinical decision aid would be premature unless these incremental-value criteria are met.

These clinical patterns provide context for the National Natural Science Foundation of China–supported investigation of Y-box binding protein 1 (YB-1), vascular smooth muscle cell senescence, and AAA progression (Grant No. 82271610), but they do not directly assess YB-1 expression or cellular senescence. Future studies linking longitudinal clinical phenotypes with vascular tissue or experimental measures of the YB-1–senescence axis may help clarify how aneurysm-specific biology relates to critical-illness trajectories.

Conclusions

Ordered laboratory measurements during the first 72 hours of ICU care identified two reproducible patterns among admissions with active AAA evidence. The multisystem derangement pattern was associated with higher mortality after the 72-hour landmark, but post hoc analyses showed substantial dependence on creatinine. The findings are best interpreted as an exploratory, renal-function-anchored description of critical illness. External validation must establish reproducibility and incremental value beyond renal and established severity measures before clinical use.

Abbreviations

AAA
abdominal aortic aneurysm
AIC
Akaike information criterion
BIC
Bayesian information criterion
CI
confidence interval
GEE
generalized estimating equations
ICD
International Classification of Diseases
ICU
intensive care unit
INR
international normalized ratio
IQR
interquartile range
MIMIC-IV
Medical Information Mart for Intensive Care IV
OR
odds ratio
RECORD
REporting of studies Conducted using Observational Routinely-collected health Data
STROBE
Strengthening the Reporting of Observational Studies in Epidemiology.

Declarations

Trial registration

Clinical trial number: not applicable.

Ethics approval and consent to participate

This study analyzed deidentified data from MIMIC-IV. The MIMIC-IV database was approved for research use by the institutional review boards of Beth Israel Deaconess Medical Center and the Massachusetts Institute of Technology, with a waiver of individual informed consent for the use of deidentified data. The investigators accessed the database after completing the required training and data-use agreement. Because this secondary analysis used only deidentified data, no additional local institutional review board approval was required.

Consent for publication

Not applicable.

Availability of data and materials

The MIMIC-IV version 3.1 data analyzed in this study are available to credentialed users through PhysioNet (https://physionet.org/content/mimiciv/3.1/; doi:10.13026/kpb9-mt58). Access requires completion of the prescribed training and acceptance of the PhysioNet data-use agreement. The authors are not permitted to redistribute patient-level MIMIC-IV data. Aggregate source data underlying the figures and tables and the Python code used to generate the reported results are provided in Supplementary File 2. Patient-level derived data remain subject to the MIMIC-IV access conditions.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 82271610). The funder had no role in study design, data analysis, interpretation, or preparation of the manuscript.

Authors' contributions

AH and FY contributed to study conception and design, data analysis, interpretation, and manuscript drafting. JL, YG, XG, ZZ, YL, WS, ZR, and ZC contributed to clinical interpretation and critical revision. YY supervised the study, obtained funding, and critically revised the manuscript. AH and FY are responsible for the statistical analysis. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the Massachusetts Institute of Technology, Beth Israel Deaconess Medical Center, and PhysioNet teams for maintaining and distributing MIMIC-IV.

Use of AI-assisted technologies

AI-assisted tools were used for language editing, code review, and document formatting. All analyses were executed locally from the supplied data. The authors reviewed and verified the code, numerical results, interpretation, references, and final text and take full responsibility for the manuscript. AI-assisted tools were not used to generate or alter the underlying scientific data.

References

1. Chaikof EL, Dalman RL, Eskandari MK, et al. The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. J Vasc Surg. 2018;67:2–e772. 10.1016/j.jvs.2017.10.044.
2. Wanhainen A, Van Herzeele I, Bastos Goncalves F, et al. European Society for Vascular Surgery 2024 clinical practice guidelines on the management of abdominal aorto-iliac artery aneurysms. Eur J Vasc Endovasc Surg. 2024;67:192–331. 10.1016/j.ejvs.2023.11.002.
3. Schermerhorn ML, O'Malley AJ, Jhaveri A, et al. Endovascular vs open repair of abdominal aortic aneurysms in the Medicare population. N Engl J Med. 2008;358:464–74. 10.1056/NEJMoa0707348.
4. Powell JT, Sweeting MJ, Ulug P, et al. Meta-analysis of individual-patient data from EVAR-1, DREAM, OVER and ACE trials comparing outcomes of endovascular or open repair for abdominal aortic aneurysm over 5 years. Br J Surg. 2017;104:166–78. 10.1002/bjs.10430.
5. Xu Z, Mao C, Su C, et al. Sepsis subphenotyping based on organ dysfunction trajectory. Crit Care. 2022;26:197. 10.1186/s13054-022-04071-4.
6. Takkavatakarn K, Oh WS, Chan L, et al. Machine learning derived serum creatinine trajectories in acute kidney injury in critically ill patients with sepsis. Crit Care. 2024;28:156. 10.1186/s13054-024-04935-x.
7. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10:1. 10.1038/s41597-022-01899-x.
8. Johnson A, Bulgarelli L, Pollard T et al. MIMIC-IV (version 3.1). PhysioNet; 2024. 10.13026/kpb9-mt58
9. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology statement: guidelines for reporting observational studies. Lancet. 2007;370:1453–7. 10.1016/S0140-6736(07)61602-X.

10. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data statement. PLoS Med. 2015;12:e1001885. 10.1371/journal.pmed.1001885.
11. Johnson A, Pollard T, Horng S, Celi LA, Mark R. MIMIC-IV-Note: deidentified free-text clinical notes (version 2.2). PhysioNet; 2023. 10.13026/1n74-ne17
12. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw. 2011;45:1–67. 10.18637/jss.v045.i03.
13. Austin PC, White IR, Lee DS, van Buuren S. Missing data in clinical research: a tutorial on multiple imputation. Can J Cardiol. 2021;37:1322–31. 10.1016/j.cjca.2020.11.010.
14. Monti S, Tamayo P, Mesirov J, Golub T. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Mach Learn. 2003;52:91–118. 10.1023/A:1023949509487.
15. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986;73:13–22. 10.1093/biomet/73.1.13.
16. Hanley JA, Negassa A, Edwardes MD, Forrester JE. Statistical analysis of correlated data using generalized estimating equations: an orientation. Am J Epidemiol. 2003;157:364–75. 10.1093/aje/kwf215.
17. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
18. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. In: Proceedings of the 9th Python in Science Conference. 2010:92–96. 10.25080/Majora-92bf1922-011
19. Wang H, Wu S, Pan D, et al. Comparison of different intensive care scoring systems and Glasgow Aneurysm score for aortic aneurysm in predicting 28-day mortality: a retrospective cohort study from MIMIC-IV database. BMC Cardiovasc Disord. 2024;24:513. 10.1186/s12872-024-04184-4.
20. Li W, Liao T, Zhang Y, Li C. Using red blood cell distribution width to predict death after abdominal aortic aneurysm rupture. BMC Cardiovasc Disord. 2023;23:172. 10.1186/s12872-023-03191-1.

Figures

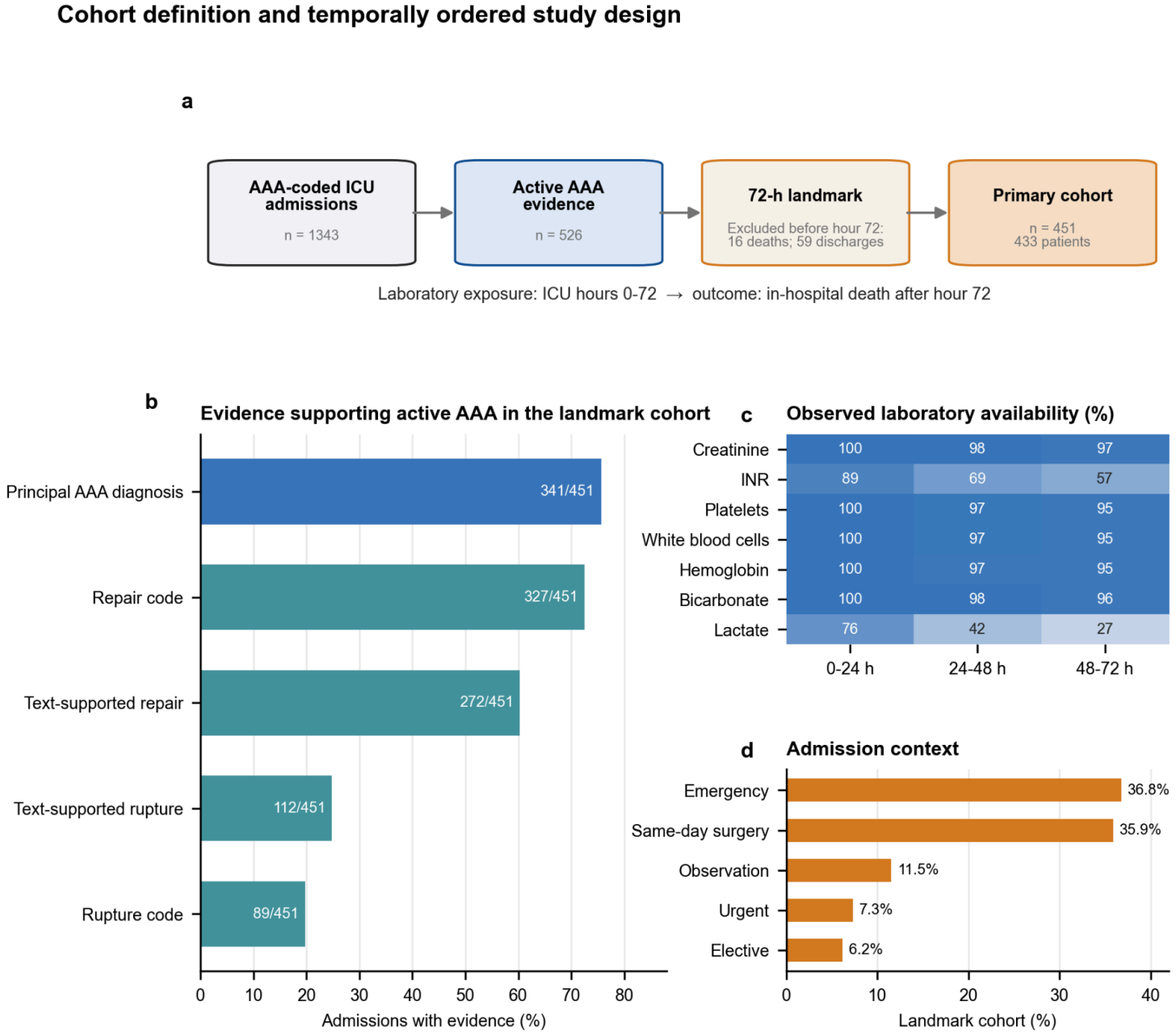


Figure 1

Cohort definition and temporally ordered study design. a Flow from 1,343 abdominal aortic aneurysm (AAA)-coded intensive care unit (ICU) admissions to the primary 72-hour landmark cohort. Active AAA evidence comprised a principal AAA diagnosis, ruptured AAA code, AAA repair code, text-supported rupture, or text-supported repair. Admissions ending in death (n = 16) or live discharge (n = 59) before hour 72 were excluded, leaving 451 admissions from 433 patients. Laboratory measurements during ICU hours 0-72 preceded ascertainment of subsequent in-hospital death. b Components supporting active AAA status; categories overlap. c Availability of laboratory measurements in each 24-hour window. Lactate was excluded from primary clustering because availability declined to 27% in hours 48-72. d Distribution of the five most frequent hospital admission types.

Longitudinal laboratory patterns and clustering stability

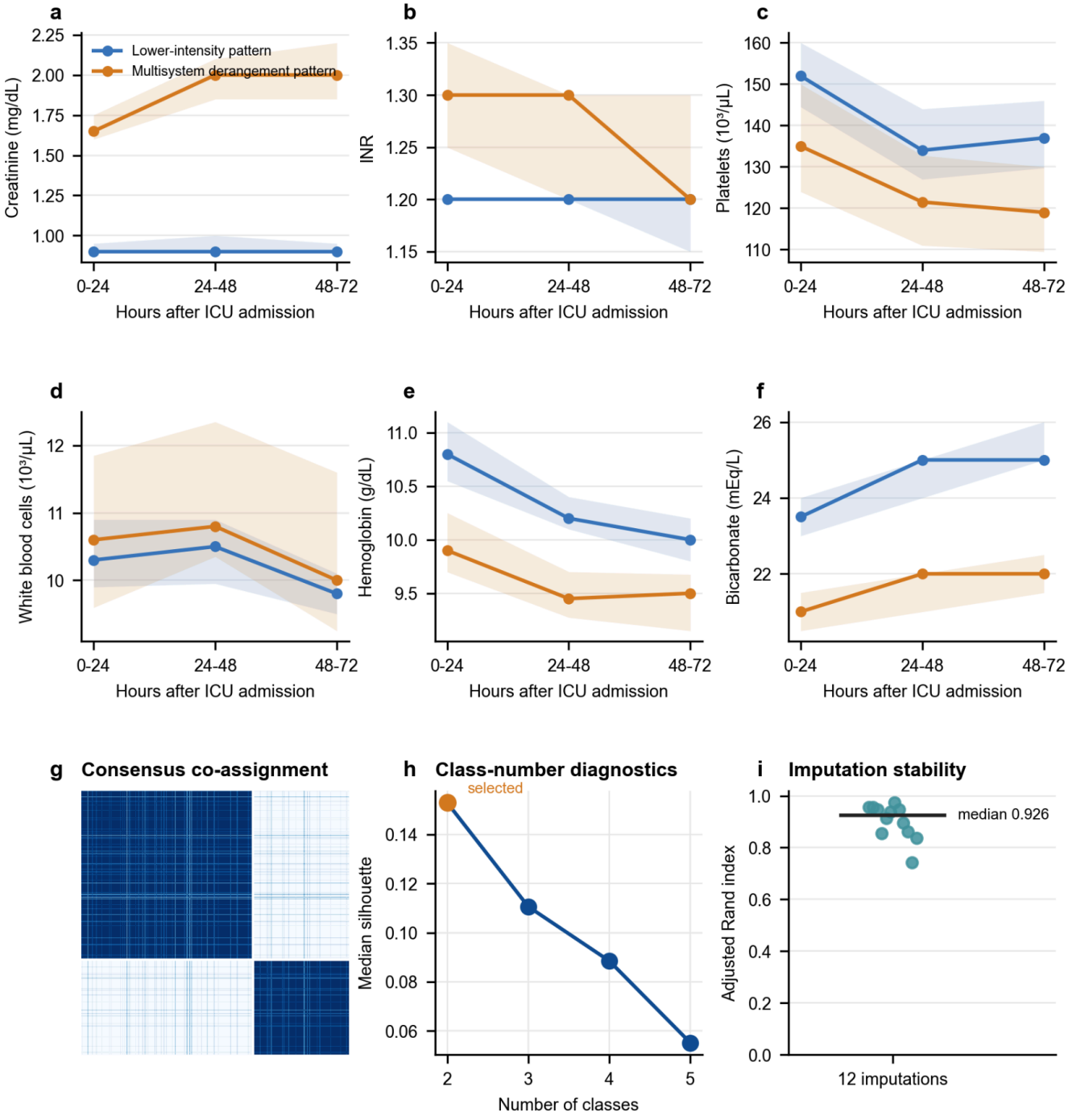


Figure 2

Longitudinal laboratory patterns and clustering stability. a-f Median creatinine, international normalized ratio (INR), platelet count, white blood cell count, hemoglobin, and bicarbonate during ICU hours 0-24, 24-48, and 48-72 in the lower-intensity pattern (n = 289) and multisystem derangement pattern (n = 162). Shading denotes nonparametric 95% confidence intervals for the observed medians. g Consensus co-assignment probabilities across 12 repeated imputations. h Median silhouette coefficient for candidate

two- to five-class models. i Adjusted Rand index comparing each imputation-specific solution with the consensus assignment; the horizontal line denotes the median (0.926).

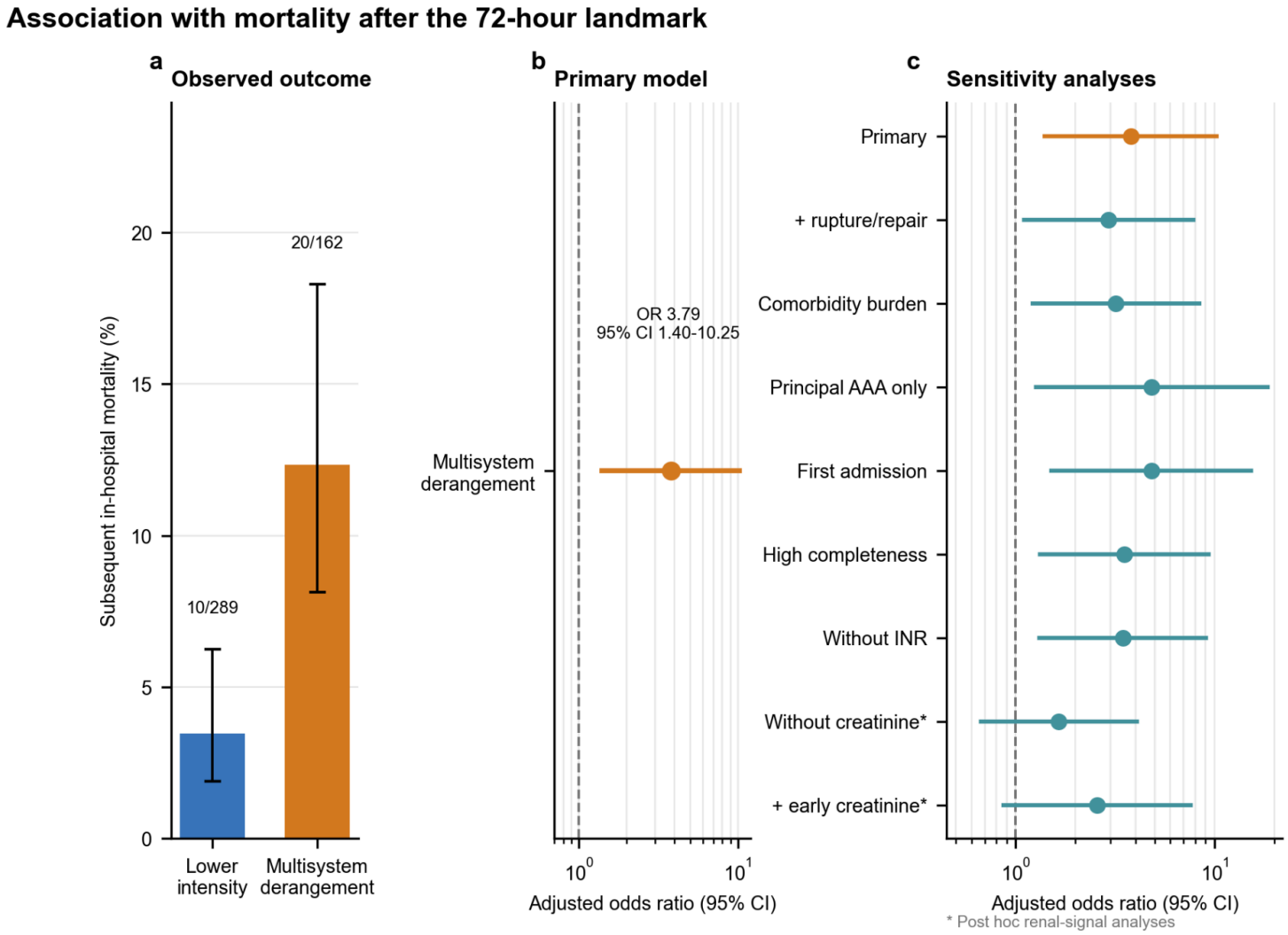


Figure 3

Association with mortality after the 72-hour landmark. a Subsequent in-hospital mortality in the lower-intensity pattern (10/289; 3.5%) and multisystem derangement pattern (20/162; 12.3%); error bars denote 95% Wilson confidence intervals. b Adjusted odds ratio comparing the multisystem derangement pattern with the lower-intensity pattern; standard errors were clustered by patient. c Primary and sensitivity estimates. The analyses excluding creatinine and additionally adjusting for early creatinine were post hoc. These estimates describe associations and do not estimate causal effects.