# **Data Dictionary – postgres.dataset.mimic\_saaki\_final (SA‑AKI Cohort)**

This document describes **all 356 columns** of the final modelling table used for SA‑AKI mortality/survival analysis. Each variable has: \* **Meaning** – plain‑language clinical definition. \* **Unit** – canonical unit in MIMIC‑IV / eICU, or *Boolean / categorical* when applicable. \* Max **2–3 explanatory lines** (brevity by design).

Because hundreds of variables share the same statistical **suffixes** (e.g. \_first, \_auc) a common legend appears first; the suffix rules apply to every laboratory, vital‑sign, and waveform base name, avoiding redundant prose.

## 1 · Suffix legend (applies to every time‑series feature)

| Suffix | Definition | Unit comment |
| --- | --- | --- |
| \_first | First measurement within the first 24 h of ICU stay. | Base unit |
| \_last | Last measurement within the same window. | Base unit |
| \_median | Median of all readings over 24 h. | Base unit |
| \_iqr | Inter‑quartile range (Q3 − Q1). | Base unit |
| \_rng | Max − Min range. | Base unit |
| \_delta | Change = last − first (directional trend). | Base unit |
| \_auc | Trapezoidal area‑under‑curve over 24 h. | Base unit × h |
| \_slope | Linear regression slope (per hour). | Base unit / h |
| \_n | Number of discrete observations. | Count |

*Every column that ends in one of these suffixes inherits its meaning from the combination:* ***base name*** *+* ***suffix meaning*** *+* ***unit given in Section 4 or 5****.*

## 2 · Identifiers & temporal anchors

| Column | Meaning | Unit |
| --- | --- | --- |
| stay\_id | Unique ICU stay identifier in MIMIC‑IV. | — |
| subject\_id | Patient identifier across admissions. | — |
| hadm\_id | Hospital admission ID (links to wards). | — |
| age | Age at ICU admission (capped ≥ 89 → 90 in MIMIC). | Years |
| gender | Recorded biological sex (M, F, or *Unknown*). | Categorical |
| ethnicity | Ethnicity string from admission note. | Categorical |
| admission\_height | Height documented on admission. | cm |
| admission\_weight | Weight documented on admission. | kg |
| bmi | Body‑mass‑index derived from height & weight. | kg / m² |
| aki\_onset\_delta\_hrs | Hours from ICU in‑time to AKI onset (KDIGO). | h |
| sepsis\_onset\_delta\_hrs | Hours from ICU in‑time to first Sepsis‑3 flag. | h |
| time\_to\_event\_hrs | Follow‑up: hours to death or last known alive. | h |
| event\_observed | **1 = in‑hospital death**, 0 = censored survival. | Boolean |

## 3 · Therapies & exposure flags

| Column | Meaning | Unit |
| --- | --- | --- |
| mechanical\_ventilation\_24hr\_flag | Any invasive ventilation in first 24 h. | Boolean |
| mechvent\_time\_to\_start\_hrs | Hours from ICU admit to first ventilation. | h |
| vasopressor\_24hr\_flag | Vasopressor infusion used in first 24 h. | Boolean |
| vasopressor\_time\_to\_start\_hrs | Hours to first vasopressor start. | h |
| rrt\_24hr\_flag | Renal‑replacement therapy within 24 h window. | Boolean |
| baseline\_scr\_estimated\_flag | Baseline creatinine was estimated (not measured). | Boolean |

## 4 · Scores & comorbidities

### 4.1 Severity scores (continuous)

| Column | Meaning | Unit |
| --- | --- | --- |
| apache\_iii\_score | Acute Physiology & Chronic Health Evaluation III. | Points |
| sofa\_total\_24hr | Total SOFA score over first day. | Points |
| sofa\_cardiovascular\_24hr | SOFA cardiovascular component. | Points |
| sofa\_coagulation\_24hr | SOFA coagulation component. | Points |
| sofa\_hepatic\_24hr | SOFA hepatic component. | Points |
| sofa\_neurological\_24hr | SOFA neurological component. | Points |
| sofa\_renal\_24hr | SOFA renal component. | Points |
| sofa\_respiratory\_24hr | SOFA respiratory component. | Points |

### 4.2 Charlson comorbidity binary flags

*Each flag = 1 if the diagnosis is present in past history / ICD codes; 0 otherwise.*

charlson\_cancer\_flag, charlson\_chf\_flag, charlson\_copd\_flag, charlson\_cvd\_flag, charlson\_dementia\_flag, charlson\_diabetes\_complications\_flag, charlson\_diabetes\_flag, charlson\_hemiplegia\_flag, charlson\_hiv\_flag, charlson\_liver\_mild\_flag, charlson\_liver\_severe\_flag, charlson\_metastatic\_cancer\_flag, charlson\_mi\_flag, charlson\_pud\_flag, charlson\_pvd\_flag, charlson\_renal\_flag, charlson\_rheumatic\_flag

## 5 · Fluid balance & urine output

| Column | Meaning | Unit |
| --- | --- | --- |
| fluid\_balance\_24hr\_ml | Net fluids (inputs − outputs) in first 24 h. | mL |
| fluid\_balance\_24hr\_ml\_perkg | Same, indexed to admission weight. | mL / kg |
| urine\_output\_total\_24hr | Total urine produced in 24 h. | mL |
| urine\_output\_total\_24hr\_perkg | Weight‑indexed urine output. | mL / kg |
| urine\_output\_avg\_hourly\_24hr | Mean hourly urine flow in window. | mL / h |

## 6 · Vital‑sign base names & units

| Base name | Clinical variable | Unit |
| --- | --- | --- |
| heartrate | Heart rate | beats / min |
| sysbp | Systolic arterial pressure | mm Hg |
| diabp | Diastolic arterial pressure | mm Hg |
| meanbp | Mean arterial pressure | mm Hg |
| resprate | Respiratory rate | breaths / min |
| spo2 | Peripheral capillary O₂ sat. | % |
| temperature | Core body temperature | °C |

*Combine each base with any suffix from Section 1 to obtain its exact column.*

## 7 · Laboratory base names & units

| Base name | What it measures | Unit |
| --- | --- | --- |
| albumin | Serum albumin | g / dL |
| alkphos | Alkaline phosphatase | U / L |
| alt | Alanine aminotransferase | U / L |
| ast | Aspartate aminotransferase | U / L |
| aniongap | Serum anion gap | mmol / L |
| bicarbonate | Serum bicarbonate | mmol / L |
| bun | Blood urea nitrogen | mg / dL |
| calcium | Total serum calcium | mg / dL |
| chloride | Serum chloride | mmol / L |
| creatinine | Serum creatinine | mg / dL |
| glucose | Plasma glucose | mg / dL |
| hematocrit | Packed‑cell volume | % |
| hemoglobin | Hemoglobin concentration | g / dL |
| inr | International Normalised Ratio | — |
| lactate | Serum lactate | mmol / L |
| lymphocytes | % lymphocytes in WBC diff | % |
| paco2 | Arterial CO₂ partial pressure | mm Hg |
| pao2 | Arterial O₂ partial pressure | mm Hg |
| ph | Arterial pH | pH units |
| plateletcount | Platelet count | 10⁹ / L |
| potassium | Serum potassium | mmol / L |
| pt | Prothrombin time | s |
| ptt | Activated partial thromboplastin time | s |
| rdw | Red‑cell distribution width | % |
| sodium | Serum sodium | mmol / L |
| totalbilirubin | Total serum bilirubin | mg / dL |
| wbc | White‑blood‑cell count | 10³ / µL |

*For every laboratory base, append any suffix to reference the exact column, e.g. creatinine\_first, bun\_auc, rdw\_slope, etc.*

## 8 · Column lookup example

**Column name:** totalbilirubin\_iqr

**Meaning:** Inter‑quartile range of total serum bilirubin values recorded in the first 24 h of ICU stay.

**Unit:** mg/dL.

The same template applies universally once you combine: \* **Base definition** (Sections 6‑7) +  \* **Suffix rule** (Section 1).

## 9 · Revision history

| Date | Notes |
| --- | --- |
| 2025‑06‑24 | Initial data dictionary generated from mimic\_saaki\_final.xlsx. |

*Maintainer:* Data Science Team – SA‑AKI Project  
*Contact:* saaki‑ds@your‑institution.edu

**End of document.**