Sepsis is a life-threatening syndrome characterized by organ dysfunction and dysregulated host response to infection. Each year, sepsis accounts for up to 20% of deaths worldwide and $20 billion of healthcare costs in the US. Sepsis and septic shock, a more severe subset, are identified using sepsis-3 criteria. Due to the large size of the time-series data, analysis was performed on a subsample of the full dataset, equalizing the number of samples in each state to 80% of the number of samples in the smallest state group. Groups identified by clustering (endotypes) were further characterized by ICU length of stay. We calculated 24 clinical and physiological features from 24 to 48 h after ICU admission. Time-series features (sampling frequency every 5 min) were extracted. Patients were categorized every 5 minutes based on the Sepsis-3 criteria, in one of four categories: No Sepsis, Suspected Sepsis [only organ dysfunction], Sepsis, or Septic Shock during their ICU stay. We calculated 24 clinical and physiological features from 24 to 48 h after ICU admission. Features were scaled and used in unsupervised dimensionality reduction and clustering algorithms, accounting for mixed data types. Due to the large size of the time-series data, analysis was performed on a subsample of the full dataset, equalizing the number of samples in each state to 80% of the number of samples in the smallest state group. Groups identified by clustering (endotypes) were further characterized by ICU length of stay and hospital discharge outcomes.

Unsupervised learning revealed four distinct clusters with distinct sepsis burdens and outcomes.

- Cluster 1 was the most severe cluster with the highest mortality and significantly longer LOS, containing all the “septic shock” patients.
- Cluster 2 contained a high proportion of sepsis and suspected sepsis patients and had intermediate mortality.
- Cluster 4 had a preponderance of “no sepsis” patients and the lowest mortality.
- Clusters 3 and 4, although comparable in terms of sepsis/septic shock burdens, were differentiable in terms of mortality (21% vs 9% respectively).

Using only time-series data, we identified four candidate endotypes associated with specific distributions of sepsis categories and specific outcome distribution. Ongoing research will further analyze identified endotypes in patient characteristics, outcomes, and response to specific treatments, as well as explore other machine learning algorithms and externally validate with other EHR data sources. In conjunction with a broader range of clinical and physiological features, these time-series-derived endotypes could help unlock the heterogeneity of sepsis, characterizing the evolution of its progression and yield greater precision in detection, prediction, and therapy, as a new paradigm for understanding and categorizing patients in the ICU who have sepsis.

References