Our Algorithm

- Our clinical state algorithm has the following features:
  - [Non-parametric]: number of clinical states is learned from the EHR data.
  - [Bias-immune]: the bias created by therapeutic intervention censoring is removed.
  - [Confidence guarantees].
- Three steps for learning:
  1. Physiological model learning.
  2. Model refinement.
  3. Domain knowledge incorporation.

The Physiological Model

- Latent variable model: clinical states \( \{z_t\}_{t \in \mathbb{N}_+} \) are hidden and manifest through lab tests and vital signs \( \{y_t\}_{t \in \mathbb{N}_+} \).
- Model for time series data \( \rightarrow \) HMM
- Non-parametric inference \( \rightarrow \) HDP prior.
- Sticky HDP-HMM with Gaussian emissions:
  - Conditional on the clinical states, the physiological variables are Gaussian.
  \[ y_t|z_t \sim \mathcal{N}(\mu(z_t), \Sigma(z_t)) \]
- To control the rate of self-transitions, we use a sticky HDP with a stick-breaking construction as a prior for the transition probabilities \( \pi_k \).

Clinical Domain Knowledge Incorporation

To assess the patients’ clinical states, attach clinical interpretation to the learned states:

- Clinicians provide labels in the EHR dataset marking specific segments of the physiological streams with clinical assessments/conditions. We use the Bhattacharyya distance to associate the discovered states with the “domain-knowledge-based” states labeled by clinicians.

This removes the bias in the learned parameters created by censoring due to interventions.

Real-World Clinical State Estimation

- We applied our algorithm to a heterogeneous cohort of 6,094 patients: admissions to Ronald Reagan UCLA medical center (March 2013-February 2016).
- Our model anticipates clinical deterioration many hours before clinicians. The model outperforms the Rothman index (currently deployed in more than 70 US hospitals).
- Clinical impact: many cardiac arrests prevented!