

Reproduced Study

Mauricio Coronado

Reproduced Study

An Interpretable Machine Learning Model for Accurate Prediction of
Sepsis in the ICU

Study Question

Can Electronic Medical Records be used to predict onset of sepsis in patients in the ICU?



Data Extraction

- Cohort exclusions depending on time.
- Extraction from multiple tables (chartevents, labevents, derived tables).
- Dataset creation for time-varying regression.
- Frequency of measurements related issues.

Methodology

- Study Approach:

Weibull-Cox proportional hazards model to predict onset of sepsis in the proceeding T hours (where $T = 4, 6, 8$ or 12)

- Reproduction Simplified Approach:

Cox's time varying proportional hazard model

$$h(t|x) = \underbrace{b_0(t)}_{\text{baseline}} \exp \left(\underbrace{\sum_{i=1}^n \beta_i (x_i(t) - \overline{x_i})}_{\text{log-partial hazard}} \right) \underbrace{\hspace{10em}}_{\text{partial hazard}}$$

Lessons Learned

- Medical databases are tricky.
- Crashes due to RAM limit exceeded!
- More detail on things like missing values imputing and description of the covariates would be helpful to improve reproduction studies.

Thank You!