

# 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:
  - Non-adults.
  - Length of ICU stay.
  - Early development of sepsis.
- Extraction from multiple sources:
  - Clinical.
  - Laboratory.
  - Demographic.
- Dataset creation for time-varying regression.

# 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

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

# Results

## All Patients

Population measure	Original Study	Reproduction
Patients (#)	42,411	34,935
Age	66	63
Male (%)	56	56
Mortality (%)	9.3	8.6
Length of Stay (Hours)	47	61
SOFA Score	1.6	-

## Septic

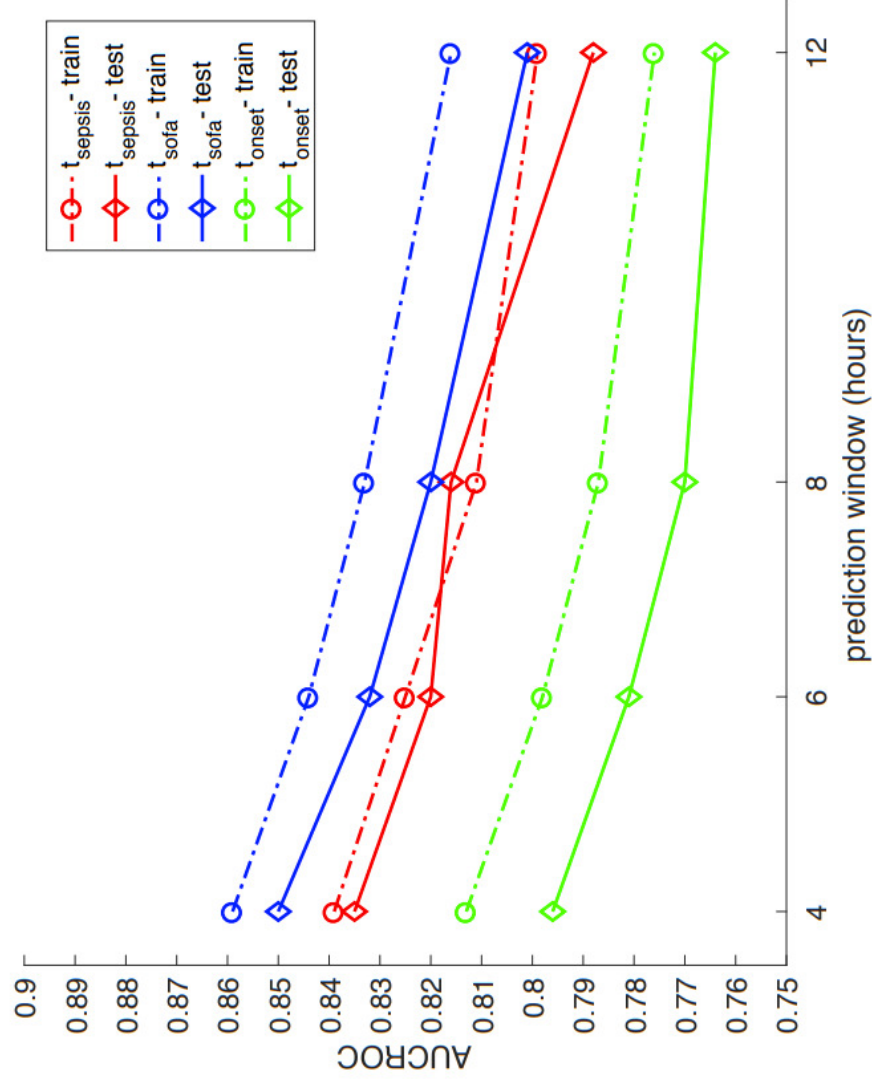
Population measure	Original Study	Reproduction
Patients (#)	3,845 (9.1%)	9,954 (28.5%)
Age	66	65*
Male (%)	59*	59*
Mortality	25*	13*
Length of Stay (Hours)	158*	94*
SOFA Score	3.3*	3.3

## Non Septic

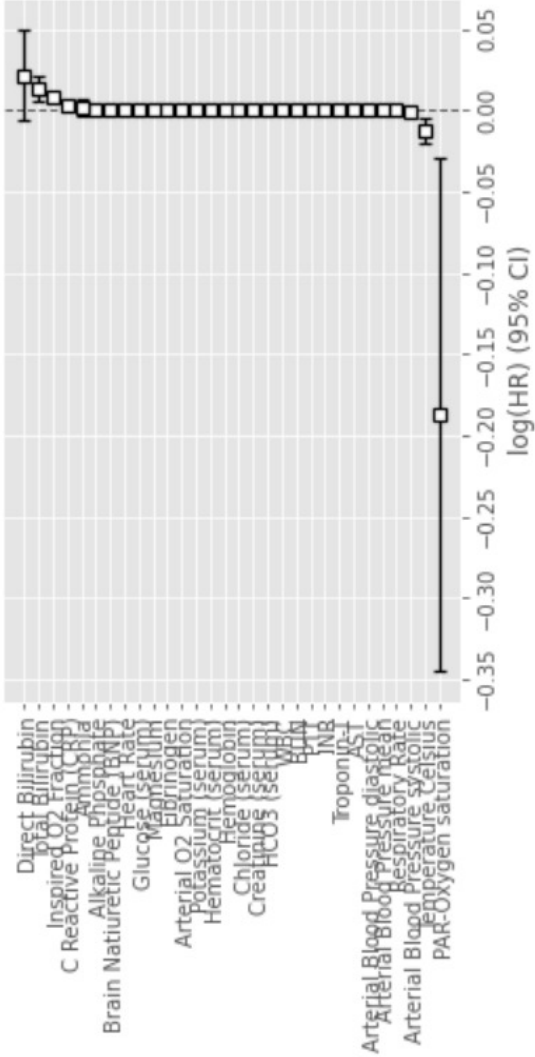
Population measure	Original Study	Reproduction
Patients (#)	38,586	24,981
Age	65	62
Male (%)	56	55
Mortality (%)	7.8	6.9
Length of Stay (Hours)	45	48
SOFA Score	1.5	-

\*Statistically significant difference between septic and non-septic patients within each cohort.

# Results



# Results



ctv.log\_likelihood\_ratio\_test()

degrees_freedom	30
null_distribution	chi squared
test_name	None
test_statistic	p
test_statistic	-log2(p)
log-likelihood ratio test	862.42
p	<0.005
-log2(p)	535.87



# Lessons Learned

- Don't underestimate medical databases.
- Efficiency in the use of resources.
- More detail on things like missing value imputing and description of the covariates would be helpful to improve reproduction studies.
- Sensitivity Analysis and a Baseline Model.

Thank You!