Learning to Detect Sepsis with a Multitask Gaussian Process RNN Classifier

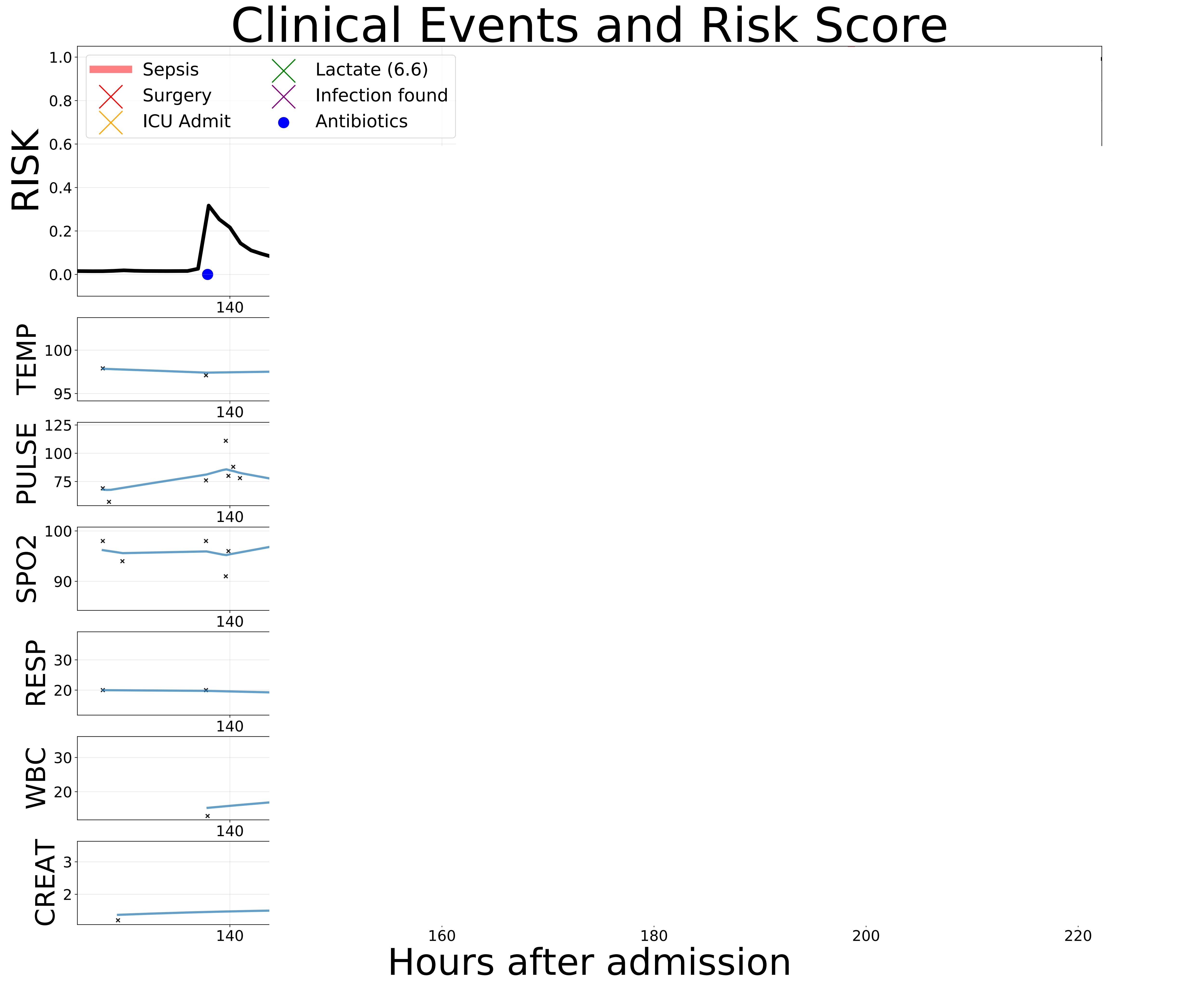
Joe Futoma
Duke University
Dept. of Statistical Science

Joint work with Sanjay Hariharan, Katherine Heller

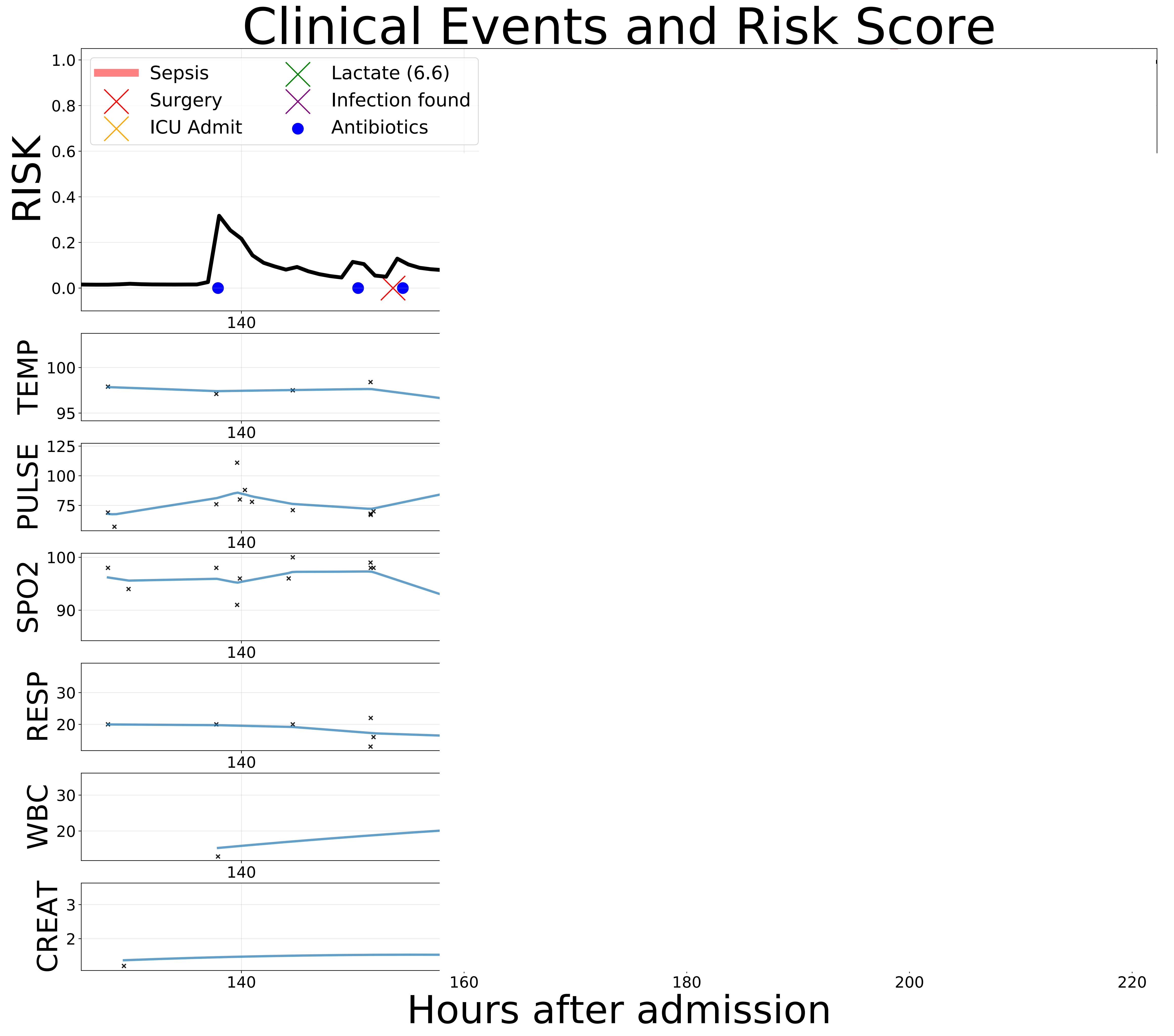
Outline

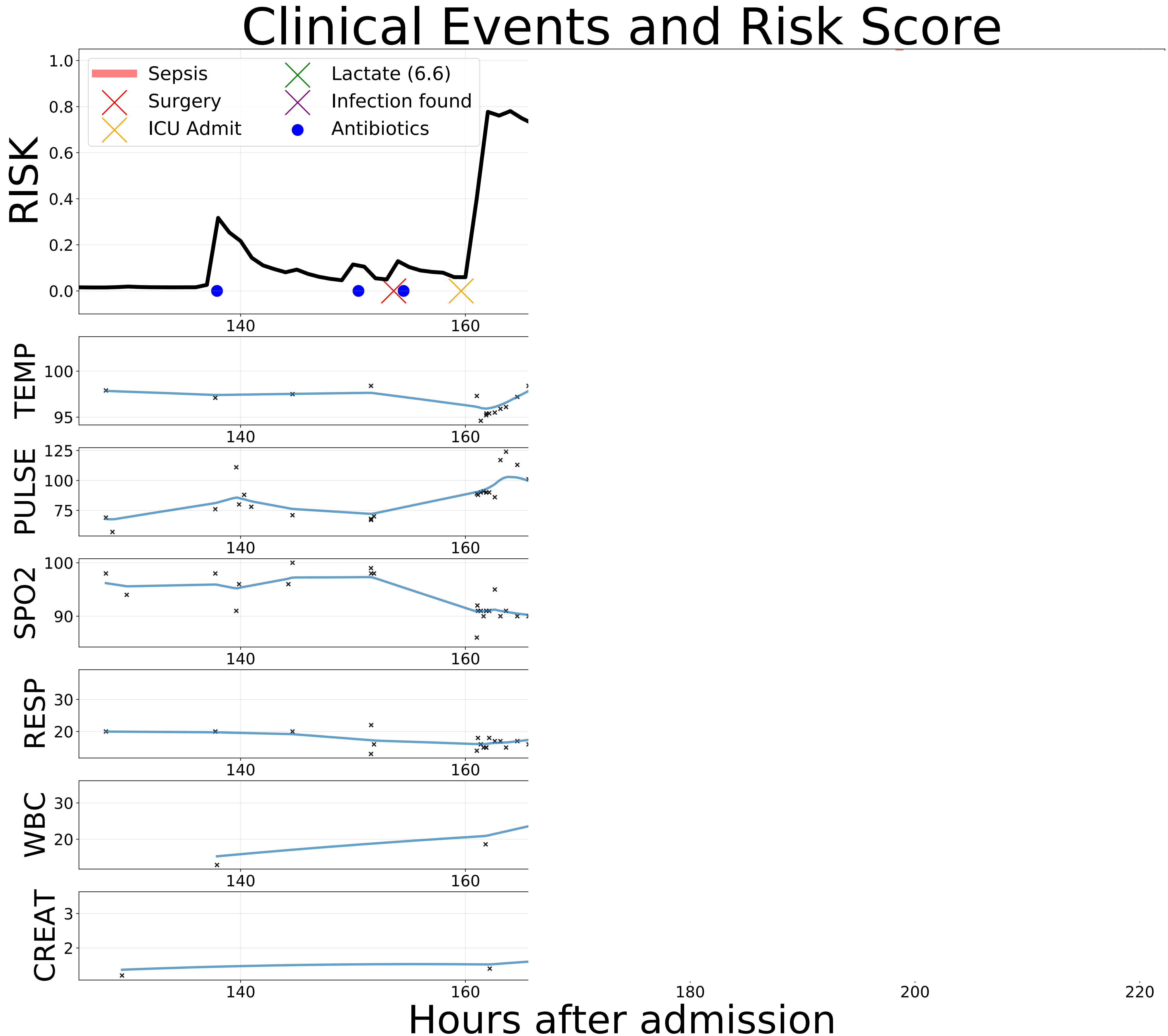
- Patient Story
- Background
- Proposed Model
- Experiments & Results
- In Clinical Practice

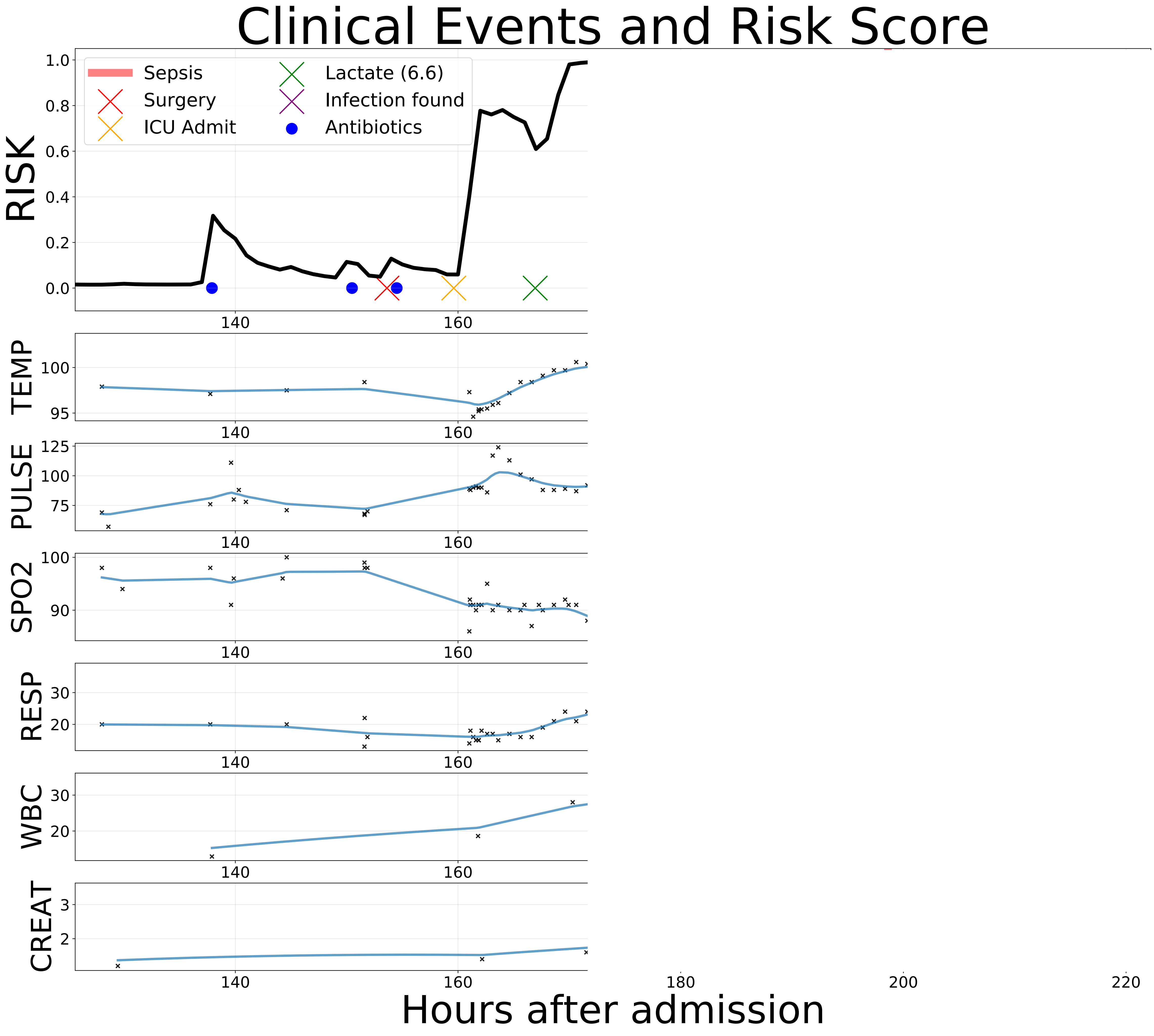
Patient Story



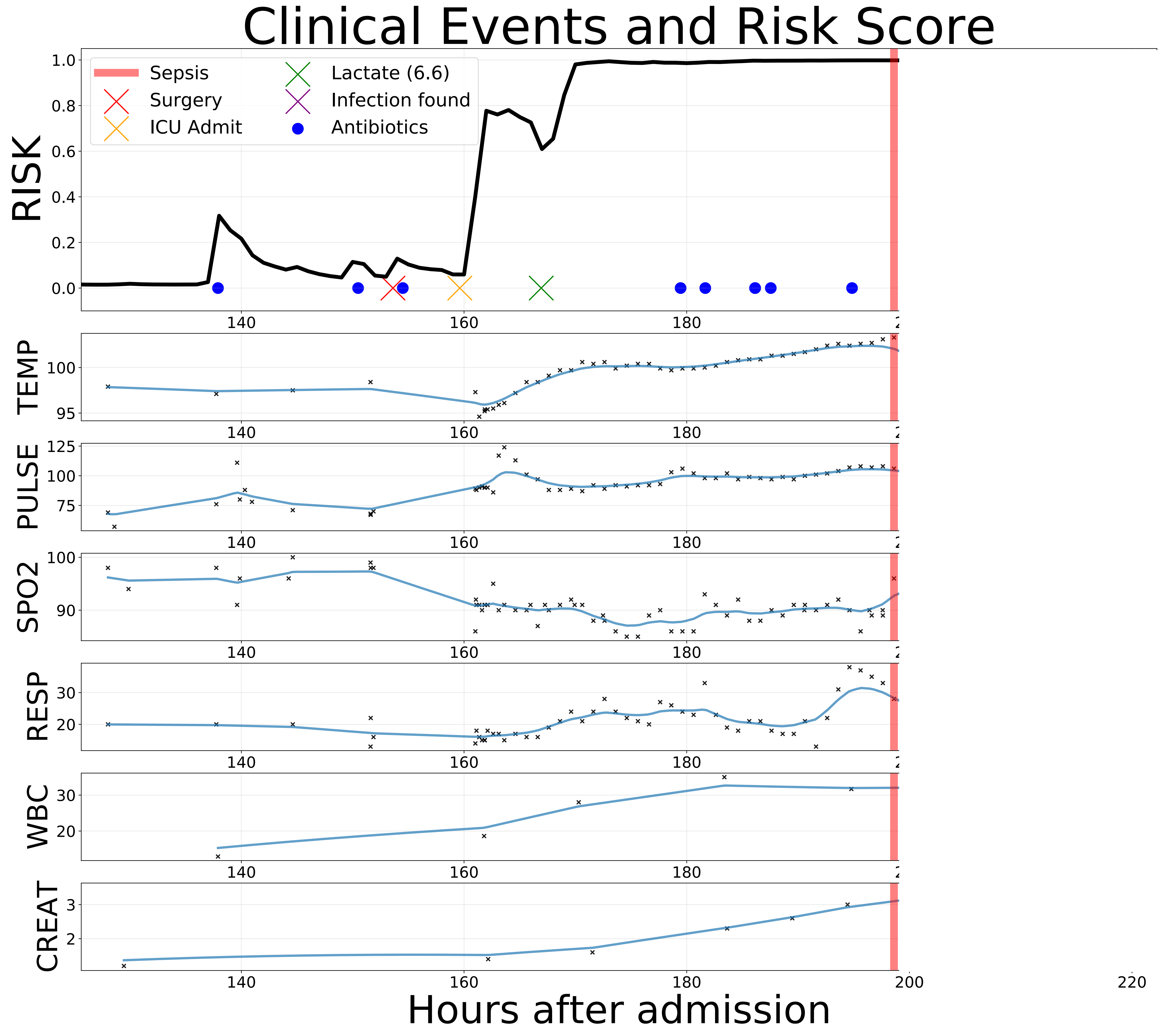
Clinical Events and Risk Score
Sepsis
Surgery
ICU Admit
Lactate (6.6)
Infection found
Antibiotics
RISK
TEMP
PULSE
SPO2
RESP
WBC
CREAT
Hours after admission


Clinical Events and Risk Score
Sepsis
Surgery
ICU Admit
Lactate (6.6)
Infection found
Antibiotics
RISK
TEMP
PULSE
SPO2
RESP
WBC
CREAT
Hours after admission

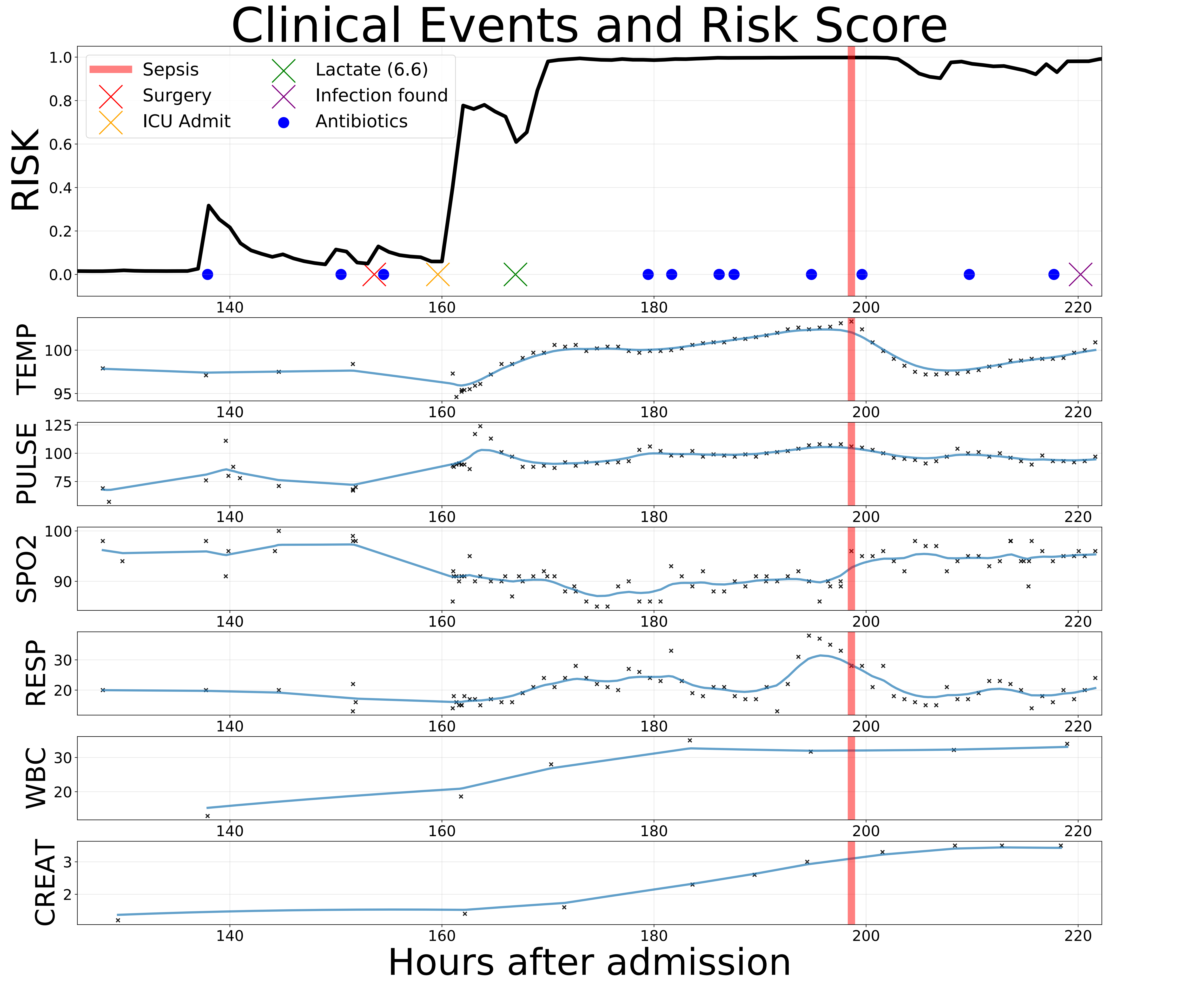

Clinical Events and Risk Score
Sepsis
Surgery
ICU Admit
Lactate (6.6)
Infection found
Antibiotics
RISK
TEMP
PULSE
SPO2
RESP
WBC
CREAT
Hours after admission



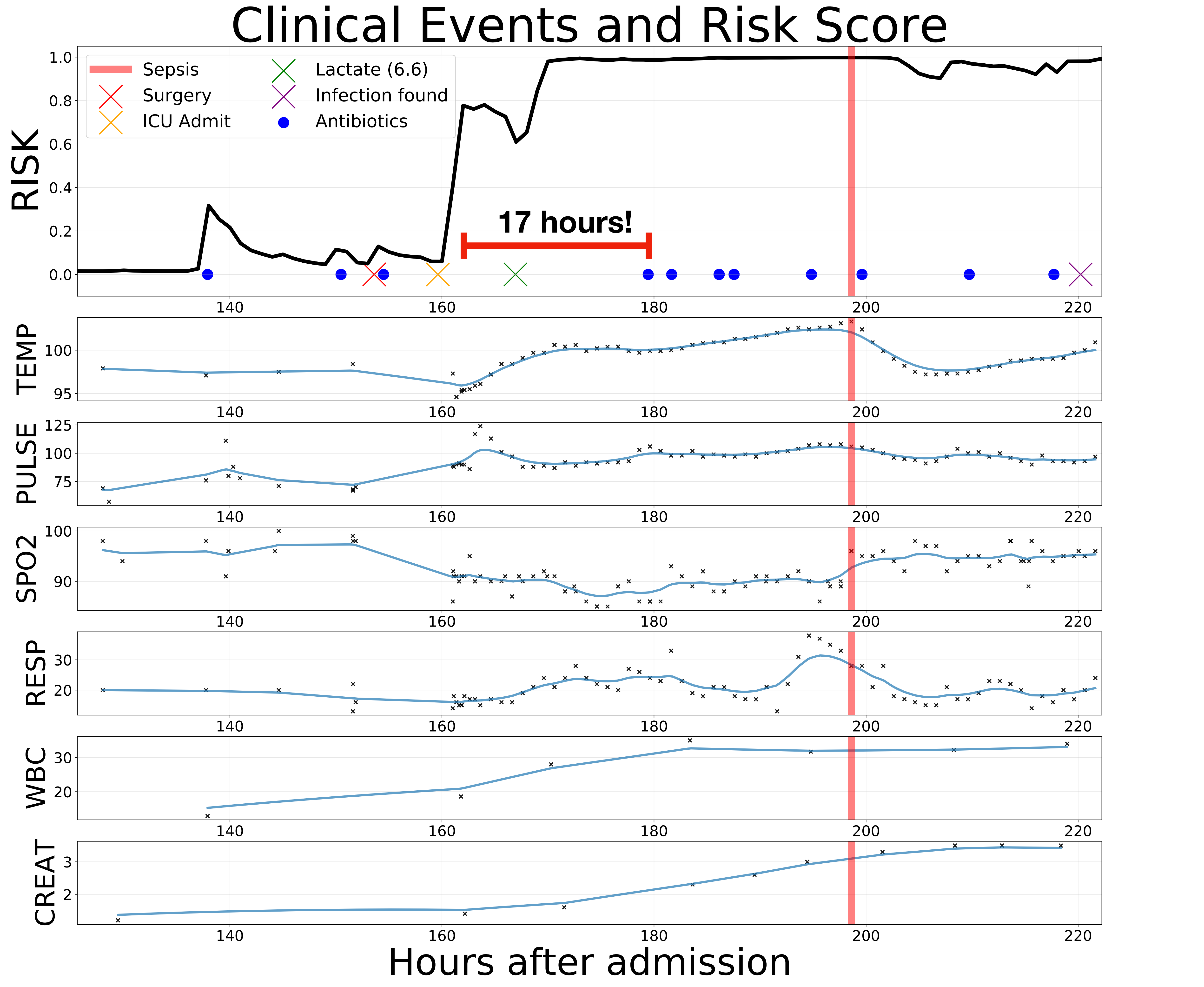
Clinical Events and Risk Score
RISK
Sepsis
Surgery
ICU Admit
Lactate (6.6)
Infection found
Antibiotics
TEMP
PULSE
SPO2
RESP
WBC
CREAT
Hours after admission

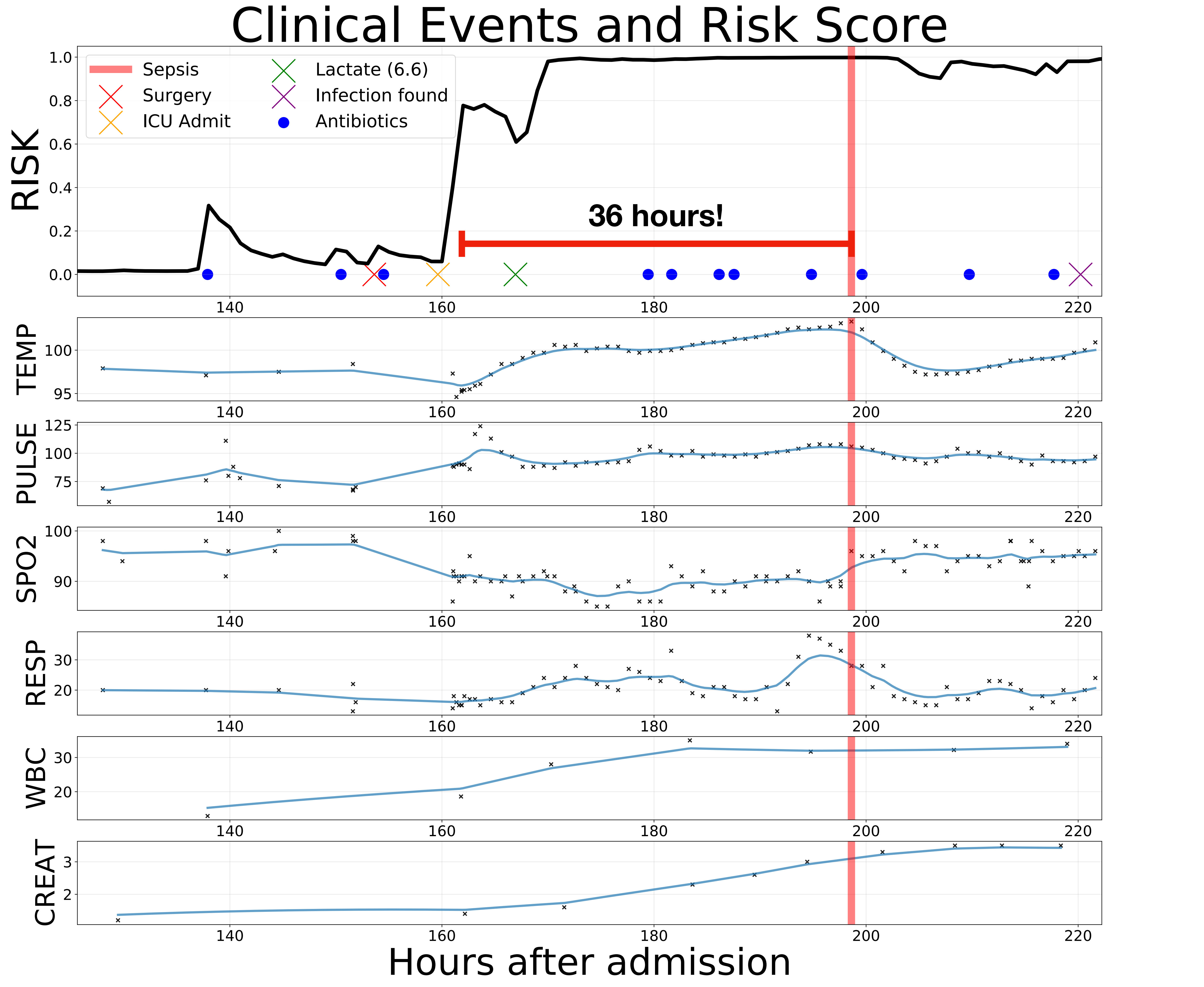


Clinical Events and Risk Score
Sepsis
Surgery
ICU Admit
Lactate (6.6)
Infection found
Antibiotics
RISK
TEMP
PULSE
SPO2
RESP
WBC
CREAT
Hours after admission



Clinical Events and Risk Score
RISK
Sepsis
Surgery
ICU Admit
Lactate (6.6)
Infection found
Antibiotics
TEMP
PULSE
SPO2
RESP
WBC
CREAT
Hours after admission


Clinical Events and Risk Score
RISK
Sepsis
Surgery
ICU Admit
Lactate (6.6)
Infection found
Antibiotics
17 hours!
TEMP
PULSE
SPO2
RESP
WBC
CREAT
Hours after admission


Clinical Events and Risk Score
RISK
Sepsis
Surgery
ICU Admit
Lactate (6.6)
Infection found
Antibiotics
36 hours!
TEMP
PULSE
SPO2
RESP
WBC
CREAT
Hours after admission

Background

Sepsis

- <u>Life-threatening complication from infection.</u>
- 750,000+ new sepsis cases each year in US; high mortality (**<u>30-50%</u>**).
- Without intervention, progress to septic shock, organ failure, death.
- **Early identification is <u>key</u>:**
    - Earlier treatment associated with improved outcomes.
- **Early identification is <u>hard</u>:**
    - No clear time of onset, no reliable biomarker (yet).

Surviving Sepsis Campaign

CMS
CENTERS FOR MEDICARE & MEDICAID SERVICES

- **<u>Sepsis Care Bundles</u>**: selected elements of care from evidence-based practice guidelines.

- In first 3 hours:

    1. Measure lactate.

    2. Get blood cultures.

    3. Give antibiotics.

- Other actions at 6 hours if no improvement.

- **We know what to do, if we know it's there!**

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Time to Treatment and Mortality during Mandated Emergency Care for Sepsis

Christopher W. Seymour, M.D., Foster Gesten, M.D., Hallie C. Prescott, M.D., Marcus E. Friedrich, M.D., Theodore J. Iwashyna, M.D., Ph.D., Gary S. Phillips, M.A.S., Stanley Lemeshow, Ph.D., Tiffany Osborn, M.D., M.P.H., Kathleen M. Terry, Ph.D., and Mitchell M. Levy, M.D

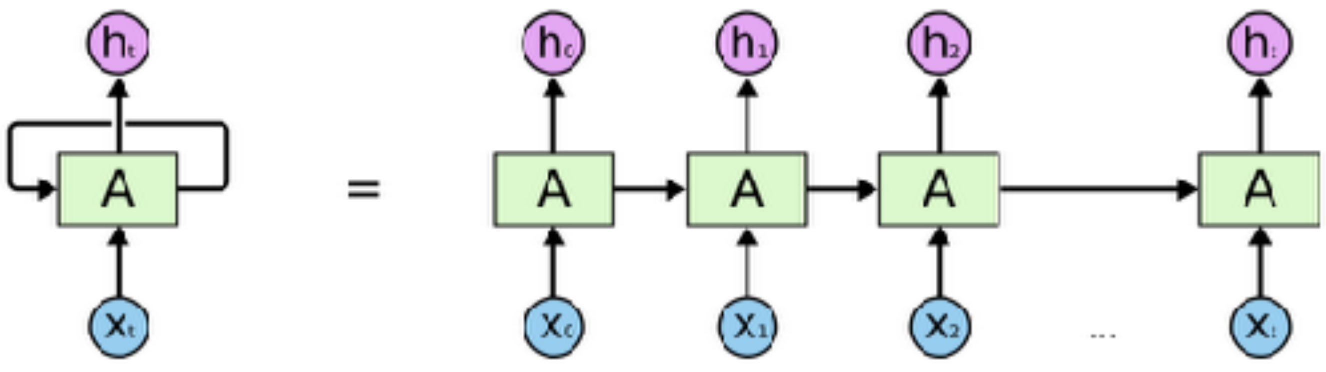
Proposed Model

Related Works

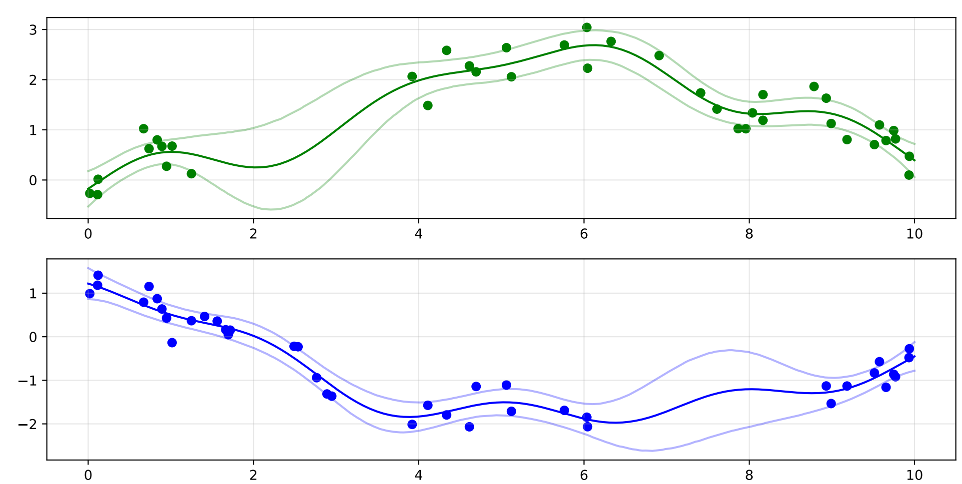
- Clinical Early Warning Scores, e.g. NEWS, SIRS, MEWS, Apache II.
    - NEWS at Duke: 63.4% of triggered alerts cancelled by nurse.
    - Typically broad, not targeted for particular conditions.
    - Low precision, leading to **high alarm fatigue**.
- (Henry et al, Science Translational Medicine 2015): TREWS score: Cox regression to predict time to septic shock, using 54 potential features [MIMIC data].
- (Ghassemi et al, AAAI 2015): Use MGPs for modeling multivariate physiological time series data from the ICU [MIMIC data].
- (Yoon et al, ICML 2016), (Hoiles & van der Schaar, NIPS 2016): related problem of predicting time to ICU admission, using streams of clinical data [UCLA in-house data].
- (Cheng-Xian & Marlin, NIPS 2016): “GP-adapter” for classifying univariate irregularly spaced time series, of the same fixed length.

Model Main Idea

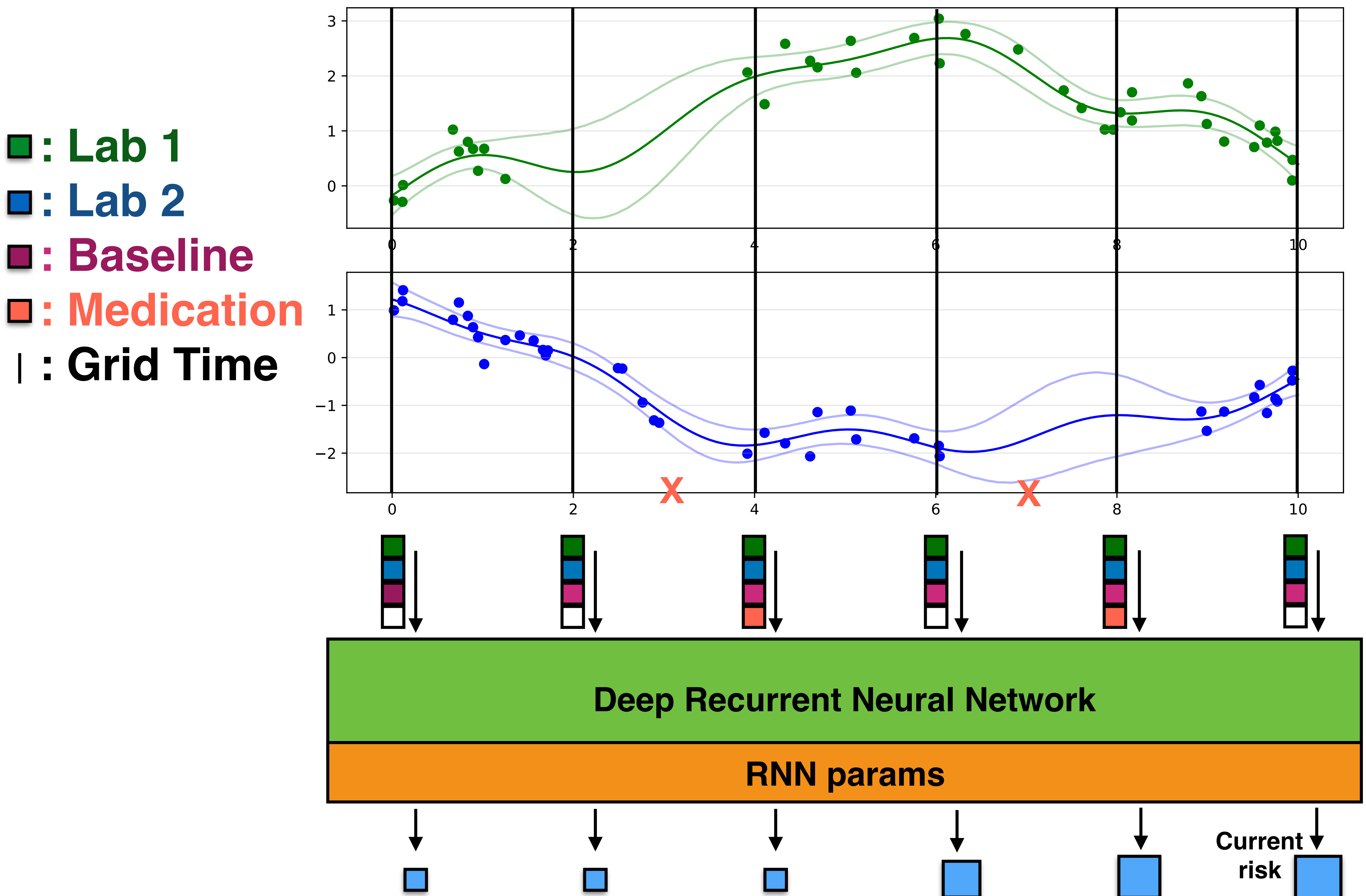
- Goal: detect onset of sepsis before it occurs.
- Data:
    1. Physiological time series (labs/vitals).
    2. Baseline admission info/comorbidities.
    3. Medication administration times.
- Multivariate time series classification: update a risk score (probability encounter is / will become septic).
- **<u>Recurrent Neural Networks (RNNs)</u>**: flexible functions, rich representational power for sequences of arbitrary length. But:
    - Requires a **complete** dataset with no missing values.
    - Requires **regularly** spaced inputs.
- **<u>Multitask Gaussian Processes</u>**: model for multivariate time series.
    - Seamlessly handles **irregularly** spaced observation times.
    - Imputes missing values on a regular grid, along with an estimate of **<u>uncertainty</u>**.



+



Model Schematic



MGP Overview

- **Gaussian process**: prior distribution over functions:

$$f_i(t) \sim \mathcal{GP}(\mu(t), K(t, t'))$$

Observation times

$$\mathbf{t}_i = (t_{i1}, \ldots, t_{iT_i})$$

T_i x T_i covariance matrix

$$f_i(\mathbf{t}_i) \sim \mathcal{N}(\mu(\mathbf{t}_i), K(\mathbf{t}_i, \mathbf{t}_i))$$

Mean function (0)

MGP Overview

- **Gaussian process**: prior distribution over functions:

$$f_i(t) \sim \mathcal{GP}(\mu(t), K(t,t'))$$

$$\mathbf{t}_i = (t_{i1}, \ldots, t_{iT_i})$$
$$f_i(\mathbf{t}_i) \sim \mathcal{N}(\mu(\mathbf{t}_i), K(\mathbf{t}_i, \mathbf{t}_i))$$

- **Multitask GP**: extension to multivariate time series.

"True" value, patient *i*, variable *m*, time *t*

M x M covariance matrix, between variables

$$k^t(t,t') = e^{-|t-t'|}$$

Correlation function, between observation times

$$\mathrm{cov}(f_{im}(t), f_{im'}(t')) = K^M_{mm'} k^t(t,t')$$

Observed value, patient *i*, variable *m*, time *t*

$$y_{im}(t) \sim \mathcal{N}(f_{im}(t), \sigma^2_m)$$

Noise level, variable *m*

MGP Overview

- **Gaussian process**: prior distribution over functions:

$$f_i(t) \sim \mathcal{GP}(\mu(t), K(t,t'))$$

$$\mathbf{t}_i = (t_{i1}, \ldots, t_{iT_i})$$
$$f_i(\mathbf{t}_i) \sim \mathcal{N}(\mu(\mathbf{t}_i), K(\mathbf{t}_i, \mathbf{t}_i))$$

- **Multitask GP**: extension to multivariate time series.

$$\mathrm{cov}(f_{im}(t), f_{im'}(t')) = K^M_{mm'} k^t(t,t')$$
$$y_{im}(t) \sim \mathcal{N}(f_{im}(t), \sigma^2_m)$$

(Completely observed)
M time series, T_i x M matrix

$$\mathrm{vec}([\mathbf{y}_{i1}, \mathbf{y}_{i2}, \ldots, \mathbf{y}_{iM}]) \equiv \mathbf{y}_i \sim \mathcal{N}(\mathbf{0}, \Sigma_i)$$
$$\Sigma_i = K^M \otimes K^{T_i} + D \otimes I,$$

T_i x T_i correlation matrix

M x M diagonal matrix, noise levels

MGP Overview

- **Gaussian process**: prior distribution over functions:

$$f_i(t) \sim \mathcal{GP}(\mu(t), K(t,t'))$$

$$\mathbf{t}_i = (t_{i1}, \dots, t_{iT_i})$$
$$f_i(\mathbf{t}_i) \sim \mathcal{N}(\mu(\mathbf{t}_i), K(\mathbf{t}_i, \mathbf{t}_i))$$

- **Multitask GP**: extension to multivariate time series.

$$\text{cov}(f_{im}(t), f_{im'}(t')) = K^M_{mm'} k^t(t,t')$$
$$y_{im}(t) \sim \mathcal{N}(f_{im}(t), \sigma^2_m)$$

$$\text{vec}([\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{iM}]) \equiv \mathbf{y}_i \sim \mathcal{N}(\mathbf{0}, \Sigma_i)$$
$$\Sigma_i = K^M \otimes K^{T_i} + D \otimes I,$$

- Define some regularly spaced (e.g. every hour) reference times, shared across all encounters.

X_i shared reference times

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iX_i})$$

$$\text{vec}(\mathbf{Z}_i) \equiv \mathbf{z}_i$$

X_i x M matrix, latent values at **x**_i

MGP Overview

- **Gaussian process**: prior distribution over functions:

$$f_i(t) \sim \mathcal{GP}(\mu(t), K(t,t'))$$

$$\mathbf{t}_i = (t_{i1}, \ldots, t_{iT_i})$$
$$f_i(\mathbf{t}_i) \sim \mathcal{N}(\mu(\mathbf{t}_i), K(\mathbf{t}_i, \mathbf{t}_i))$$

- **Multitask GP**: extension to multivariate time series.

$$\mathrm{cov}(f_{im}(t), f_{im'}(t')) = K^M_{mm'} k^t(t,t')$$
$$y_{im}(t) \sim \mathcal{N}(f_{im}(t), \sigma^2_m)$$

$$\mathrm{vec}([\mathbf{y}_{i1}, \mathbf{y}_{i2}, \ldots, \mathbf{y}_{iM}]) \equiv \mathbf{y}_i \sim \mathcal{N}(\mathbf{0}, \Sigma_i)$$
$$\Sigma_i = K^M \otimes K^{T_i} + D \otimes I,$$

- Define some regularly spaced (e.g. every hour) reference times, shared across all encounters.

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iX_i})$$

$$\mathrm{vec}(\mathbf{Z}_i) \equiv \mathbf{z}_i$$

Given **y**_i, conditional normal posterior:

X_i x T_i correlation matrix

$$\mu_{z_i} = (K^M \otimes K^{X_i T_i})\Sigma_i^{-1}\mathbf{y}_i$$
$$\Sigma_{z_i} = (K^M \otimes K^{X_i}) - (K^M \otimes K^{X_i T_i})\Sigma_i^{-1}(K^M \otimes K^{T_i X_i})$$

X_i x X_i correlation matrix

- MGP posterior for Z_i: the M labs at X_i times; maintaining uncertainty.

MGP Overview

- **Gaussian process**: prior distribution over functions:

$$f_i(t) \sim \mathcal{GP}(\mu(t), K(t,t'))$$

$$\mathbf{t}_i = (t_{i1}, \ldots, t_{iT_i})$$
$$f_i(\mathbf{t}_i) \sim \mathcal{N}(\mu(\mathbf{t}_i), K(\mathbf{t}_i, \mathbf{t}_i))$$

- **Multitask GP**: extension to multivariate time series.

$$\mathrm{cov}(f_{im}(t), f_{im'}(t')) = K^M_{mm'} k^t(t,t')$$
$$y_{im}(t) \sim \mathcal{N}(f_{im}(t), \sigma^2_m)$$

$$\mathrm{vec}([\mathbf{y}_{i1}, \mathbf{y}_{i2}, \ldots, \mathbf{y}_{iM}]) \equiv \mathbf{y}_i \sim \mathcal{N}(\mathbf{0}, \Sigma_i)$$
$$\Sigma_i = K^M \otimes K^{T_i} + D \otimes I,$$

- Define some regularly spaced (e.g. every hour) reference times, shared across all encounters.

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iX_i})$$

$$\mathrm{vec}(\mathbf{Z}_i) \equiv \mathbf{z}_i$$

MGP parameters to learn, shared across all encounters

$$\boxed{\theta = (K^M, \{\sigma^2_m\}^M_{m=1}, l)}$$

$$\mu_{z_i} = (K^M \otimes K^{X_i T_i}) \Sigma_i^{-1} \mathbf{y}_i$$
$$\Sigma_{z_i} = (K^M \otimes K^{X_i}) - (K^M \otimes K^{X_i T_i}) \Sigma_i^{-1} (K^M \otimes K^{T_i X_i})$$

- MGP posterior for Z_i: the M labs at X_i times; maintaining uncertainty.

MGP-RNN

- RNN input: latent values Z_i, baseline covariates, medication indicators.

$$\mathbf{d}_{ij} = [\mathbf{z}_{ij}^\top, \mathbf{b}_i^\top, \mathbf{p}_{ij}^\top]^\top$$

Latent series values

Baseline Covariates

Medications Indicators

MGP-RNN

- RNN input: latent values Z_i, baseline covariates, medication indicators. $$\mathbf{d}_{ij} = [\mathbf{z}_{ij}^{\top}, \mathbf{b}_i^{\top}, \mathbf{p}_{ij}^{\top}]^{\top}$$

- To learn RNN parameters: optimize loss comparing model predictions to true label.

RNN prediction

label

Loss $l(f(\mathbf{D}_i; \mathbf{w}), o_i)$

RNN input matrix

RNN Parameters

MGP-RNN

- RNN input: latent values Z_i, baseline covariates, medication indicators. $\mathbf{d}_{ij} = [\mathbf{z}_{ij}^\top, \mathbf{b}_i^\top, \mathbf{p}_{ij}^\top]^\top$

- To learn RNN parameters: optimize loss comparing model predictions to true label. $l(f(\mathbf{D}_i; \mathbf{w}), o_i)$

- **<u>Problem</u>**: **z**_ij are not observed!

MGP-RNN

- RNN input: latent values Z_i, baseline covariates, medication indicators. $\mathbf{d}_{ij} = [\mathbf{z}_{ij}^\top, \mathbf{b}_i^\top, \mathbf{p}_{ij}^\top]^\top$

- To learn RNN parameters: optimize loss comparing model predictions to true label. $l(f(\mathbf{D}_i; \mathbf{w}), o_i)$

- **<u>Problem</u>**: **z**_ij are not observed!

- Instead, optimize expected loss with respect to MGP posterior on **z**_i. Overall learning problem:

$$\boxed{\mathbf{w}^*, \theta^* = \mathrm{argmin}_{w,\theta} \sum_{i=1}^{N} \mathbb{E}_{z_i \sim N(\mu_{z_i}, \Sigma_{z_i}; \theta)}[l(f(\mathbf{D}_i; \mathbf{w}), o_i)]}$$

MGP-RNN

- RNN input: latent values Z_i, baseline covariates, medication indicators. $\mathbf{d}_{ij} = [\mathbf{z}_{ij}^\top, \mathbf{b}_i^\top, \mathbf{p}_{ij}^\top]^\top$

- To learn RNN parameters: optimize loss comparing model predictions to true label. $l(f(\mathbf{D}_i; \mathbf{w}), o_i)$

- **<u>Problem</u>**: **z**_ij are not observed!

- Instead, optimize expected loss with respect to MGP posterior on **z**_i. Overall learning problem:
  - End-to-end learning (jointly learn MGP, RNN). $$\boxed{\mathbf{w}^*, \theta^* = \operatorname{argmin}_{w,\theta} \sum_{i=1}^{N} \mathbb{E}_{z_i \sim N(\mu_{z_i}, \Sigma_{z_i}; \theta)}[l(f(\mathbf{D}_i; \mathbf{w}), o_i)]}$$
  - Reparameterization trick to get gradients, MC to approximate expectation.
  - Optimize with stochastic gradient descent.
  - Conjugate gradient, Lanczos method to speed computation.

MGP-RNN

- RNN input: latent values Z_i, baseline covariates, medication indicators. $\mathbf{d}_{ij} = [\mathbf{z}_{ij}^\top, \mathbf{b}_i^\top, \mathbf{p}_{ij}^\top]^\top$

- To learn RNN parameters: optimize loss comparing model predictions to true label.

$$l(f(\mathbf{D}_i; \mathbf{w}), o_i)$$

- **Problem**: **z**_ij are not observed!

- Instead, optimize expected loss with respect to MGP posterior on **z**_i. Overall learning problem:

  - End-to-end learning (jointly learn MGP, RNN).

$$\boxed{\mathbf{w}^*, \theta^* = \mathrm{argmin}_{w,\theta} \sum_{i=1}^{N} \mathbb{E}_{z_i \sim N(\mu_{z_i}, \Sigma_{z_i}; \theta)} [l(f(\mathbf{D}_i; \mathbf{w}), o_i)]}$$

  - Reparameterization trick to get gradients, MC to approximate expectation.

Risk score for new patient *i'*

$$\boxed{\mathbb{E}_{z_{i'} \sim N(\mu_{z_{i'}}, \Sigma_{z_{i'}}; \theta^*)} [f(\mathbf{D}_{i'}; \mathbf{w}^*)]}$$

  - Optimize with stochastic gradient descent.
  - Conjugate gradient, Lanczos method to speed computation.

Experiments & Results

Dataset

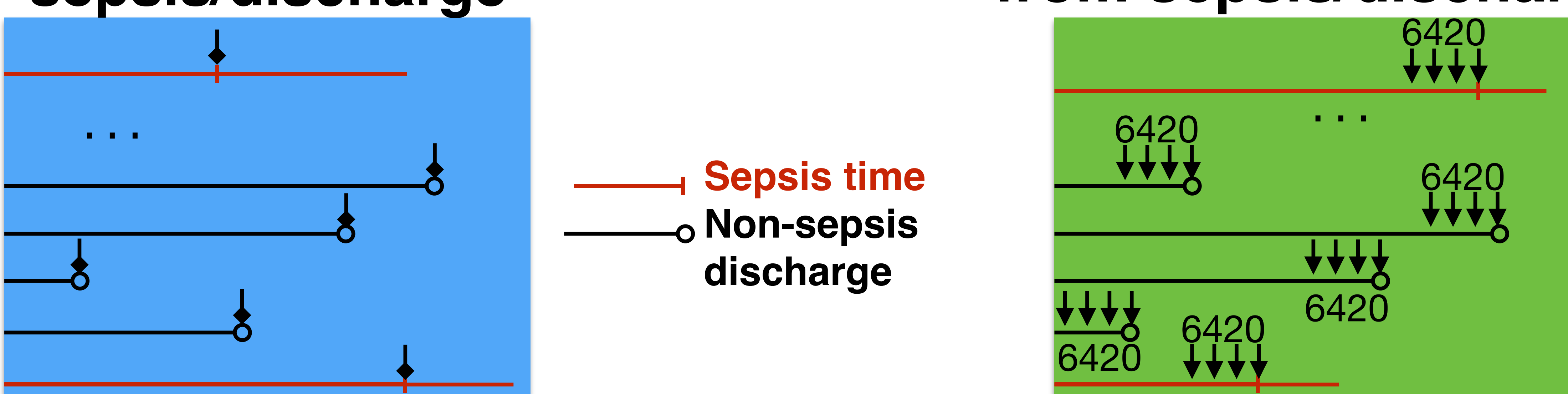
- **49,312** inpatient encounters at Duke Hospital over 18 months, 21.4% with a sepsis event; no specific inclusion/exclusion criteria.
- **34** physiological variables (5 vitals, 29 labs).
    - At least one value for each vital in 99% of encounters.
    - Some labs rarely measured (2-4%), most measured 20-80% of the time.
- **35** baseline covariates (e.g. age, transfer status, comorbidities).
- **8** medication classes (e.g. antibiotics, opioids, heparins).
- Mean length of stay 121.7 hours (sd: 108.1); highly variable.

Experimental Setup

- Compare metrics hours in advance of sepsis/discharge:
  - **AU-ROC:** Area under ROC curve / C-statistic.
  - **AU-PR:** Area under Precision/Recall curve.
  - **Precision:** At a fixed sensitivity (0.85).

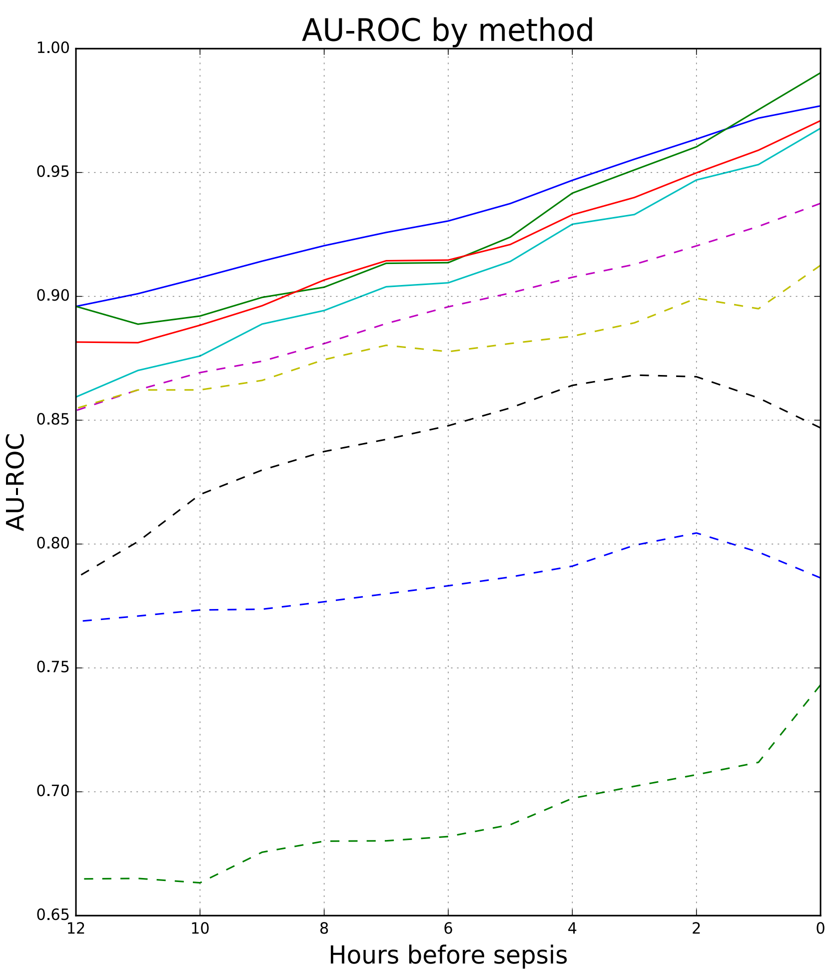
Train: up to sepsis/discharge

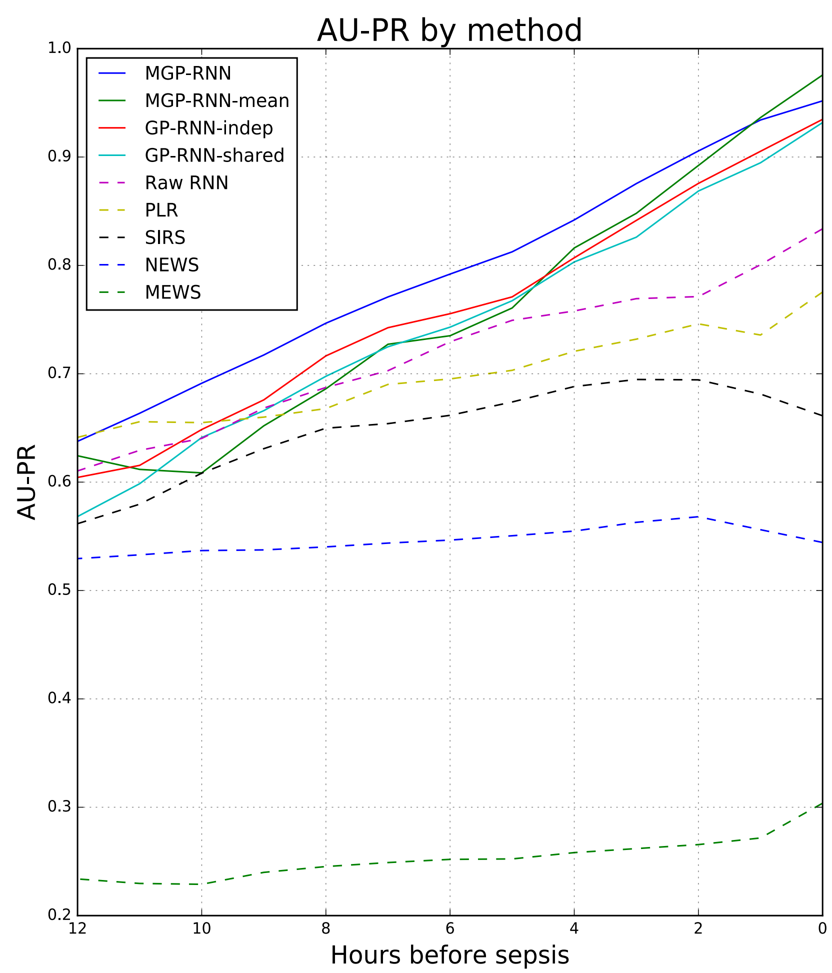
Test: vary # hours from sepsis/discharge

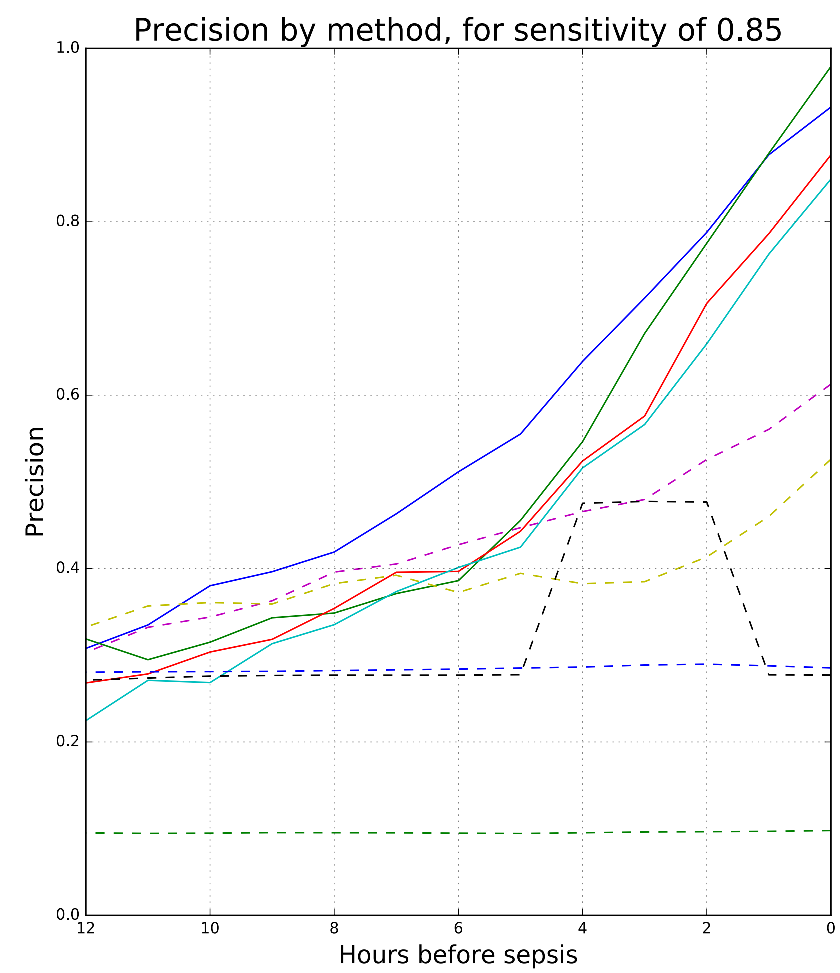


Results

- **MGP-RNN**, **MGP-RNN-mean**, **GP-RNN-indep**, **GP-RNN-shared:** variations on our approach [solid colors]
- **Raw RNN**: trained on raw data with no GP (missing: carry forwards last observed value)
- **PLR**: Penalized logistic regression, same imputation as Raw RNN
- **SIRS**, **NEWS**, **MEWS**: clinical scores





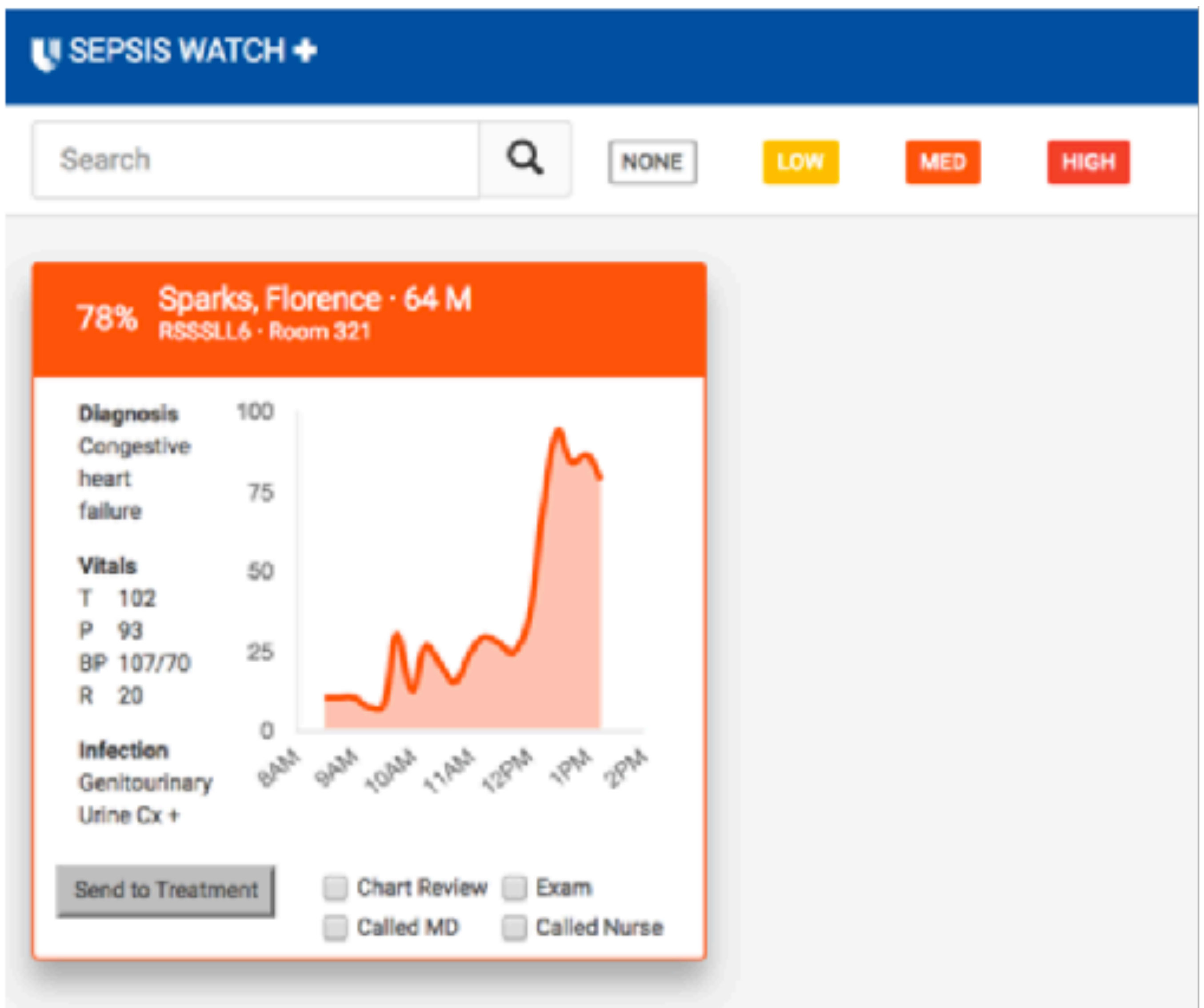


In Clinical Practice

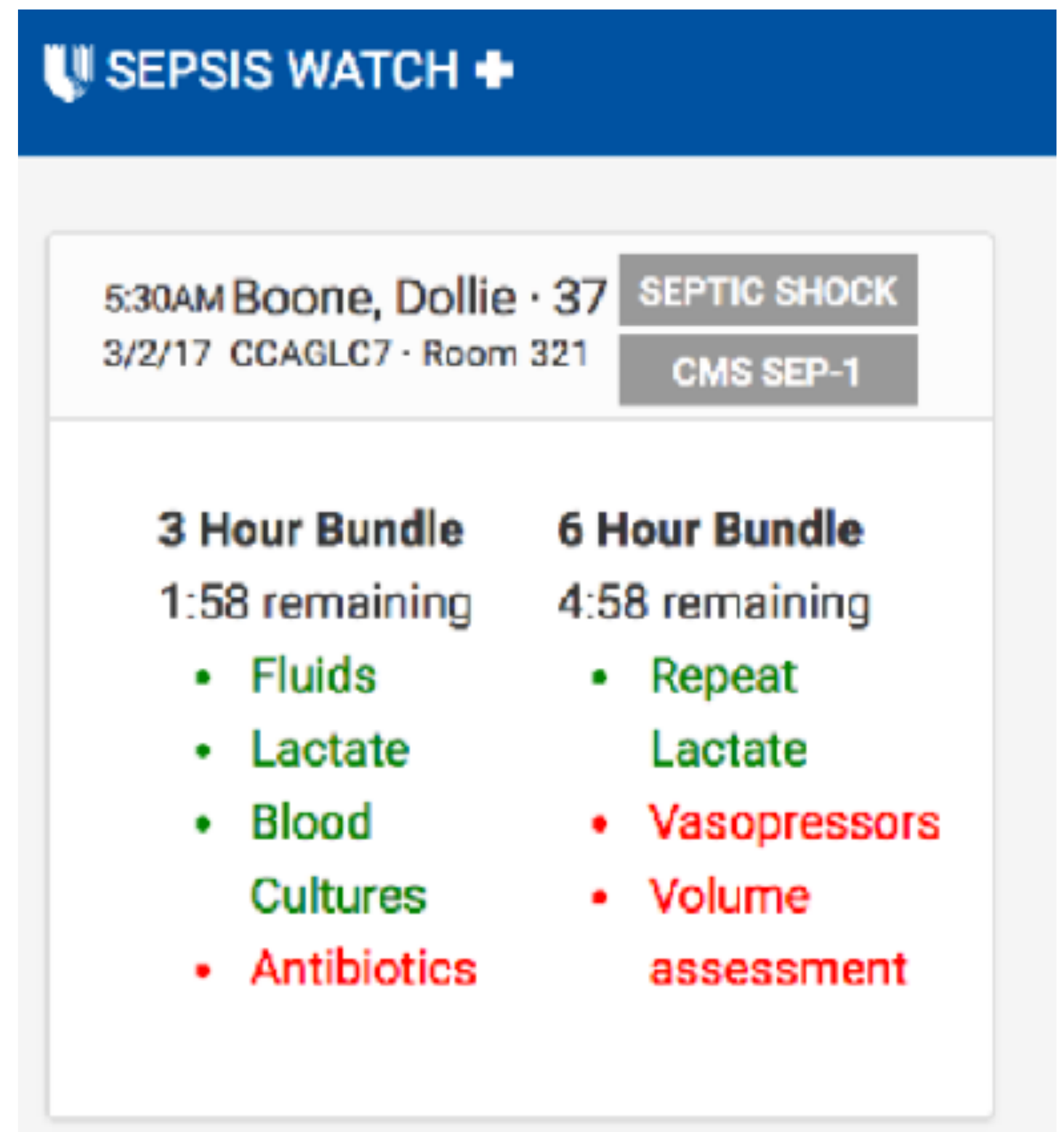
SepsisWatch

SepsisWatch

Watchlist



Treatment



…but does it work? Let’s find out!

- **Sepsis Rapid Response Team** (RRT): Fast-responding team to help coordinate care for patients with suspected sepsis.
  - Cardiac care unit nurses, pharmacists, hospitalists, respiratory therapists, administrators (logistics).
- Patients at high risk, or that meet the sepsis definition, will be reviewed by a care nurse.
- In planning stages of **randomized clinical trial!**
  - Goal: evaluate effect of using the app on **clinical outcomes** (in-hospital mortality).
  - Secondary outcomes: compliance to completing 3, 6 hour bundles on time.
  - On target to launch **this fall!**

Conclusion

- Novel model for early detection of sepsis, leveraging **deep learning** and **Gaussian processes.**
- Significantly improved performance over NEWS used at Duke.
- To be used in actual practice in a **clinical trial!**
- Many exciting new directions:
  - Model other clinical events (e.g. code blue).
  - RNN Attention mechanism for interpretability.
  - Learn treatment-response curves from medications data.
  - Reinforcement Learning to recommend optimal treatments.

Acknowledgements

DIHI Team:
Mark Sendak, MD, MPP
Nathan Brajer, MD Candidate
Michael Gao
Suresh Balu, MBA
Machine Learning:
Sanjay Hariharan, MS
Katherine Heller, PhD
Sepsis Clinicians:
Cara O'Brien, MD
Armando Bedoya, MD
Meredith Clement, MD
Software Developer:
Faraz Yashar

Duke Statistical Science

NDSEG
NATIONAL DEFENSE SCIENCE & ENGINEERING GRADUATE FELLOWSHIP

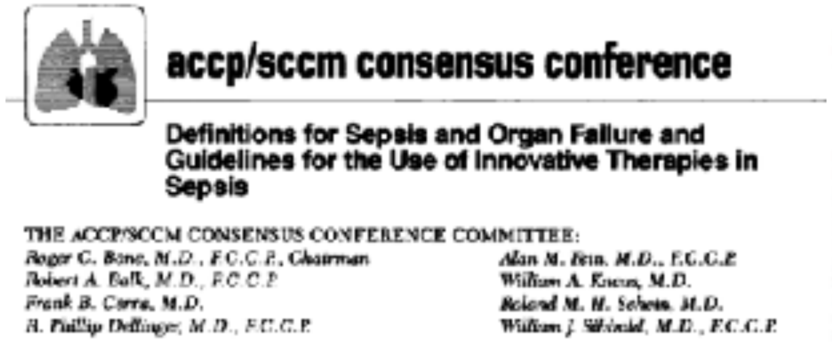
Duke Institute for Health Innovation

jdf38@duke.edu

https://github.com/jfutoma/MGP-RNN

Poster 43

Defining Sepsis

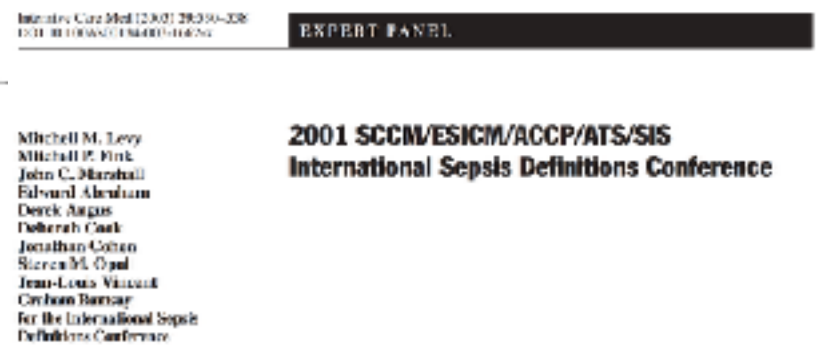

accp/sccm consensus conference

Definitions for Sepsis and Organ Failure and Guidelines for the Use of Innovative Therapies in Sepsis

THE ACCP/SCCM CONSENSUS CONFERENCE COMMITTEE:

Roger C. Bone, M.D., F.C.C.P., Chairman
Robert A. Balk, M.D., F.C.C.P.
Frank B. Cerra, M.D.
R. Phillip Dellinger, M.D., F.C.C.P.
Alan M. Fein, M.D., F.C.C.P.
William A. Knaus, M.D.
Roland M. H. Schein, M.D.
William J. Sibbald, M.D., F.C.C.P.

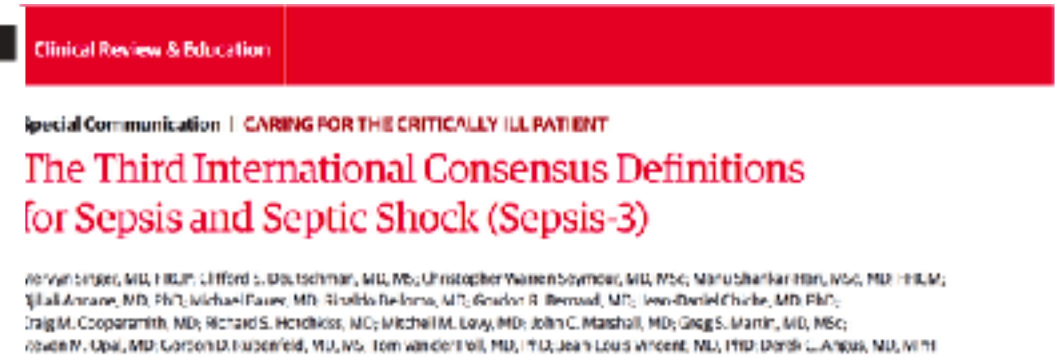
1992


EXPERT PANEL

2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference

Mitchell M. Levy
Mitchell P. Fink
John C. Marshall
Edward Abraham
Derek Angus
Deborah Cook
Jonathan Cohen
Steven M. Opal
Jean-Louis Vincent
Graham Ramsay
for the International Sepsis Definitions Conference

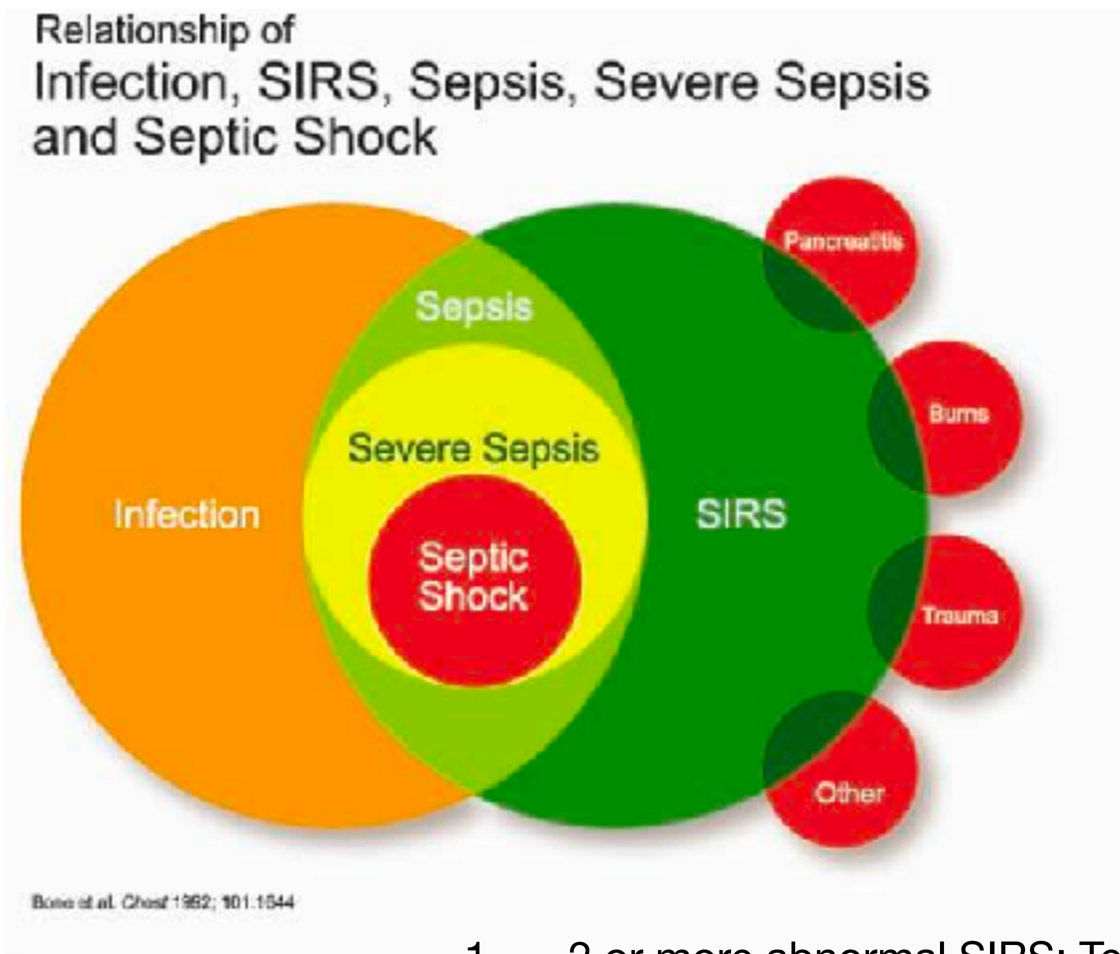
2001

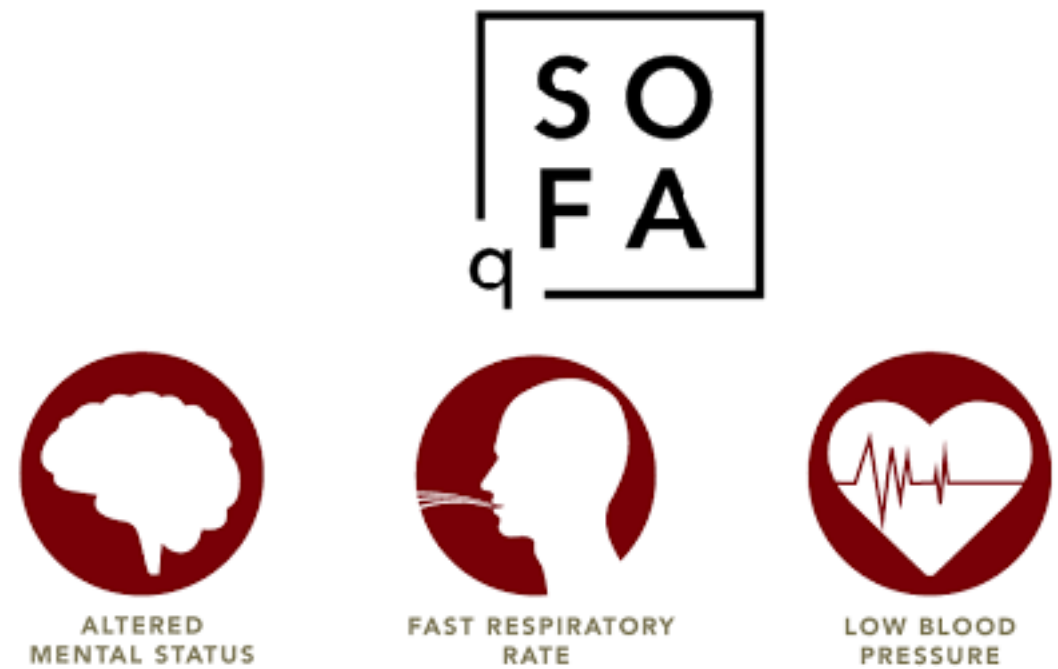

Clinical Review & Education

Special Communication | CARING FOR THE CRITICALLY ILL PATIENT

The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)

2016





Our definition ("Severe Sepsis")

1. 2 or more abnormal SIRS: Temperature, Heart Rate, Respiration Rate, WBC Count.
2. Blood culture (suspected infection).
3. End organ damage lab.

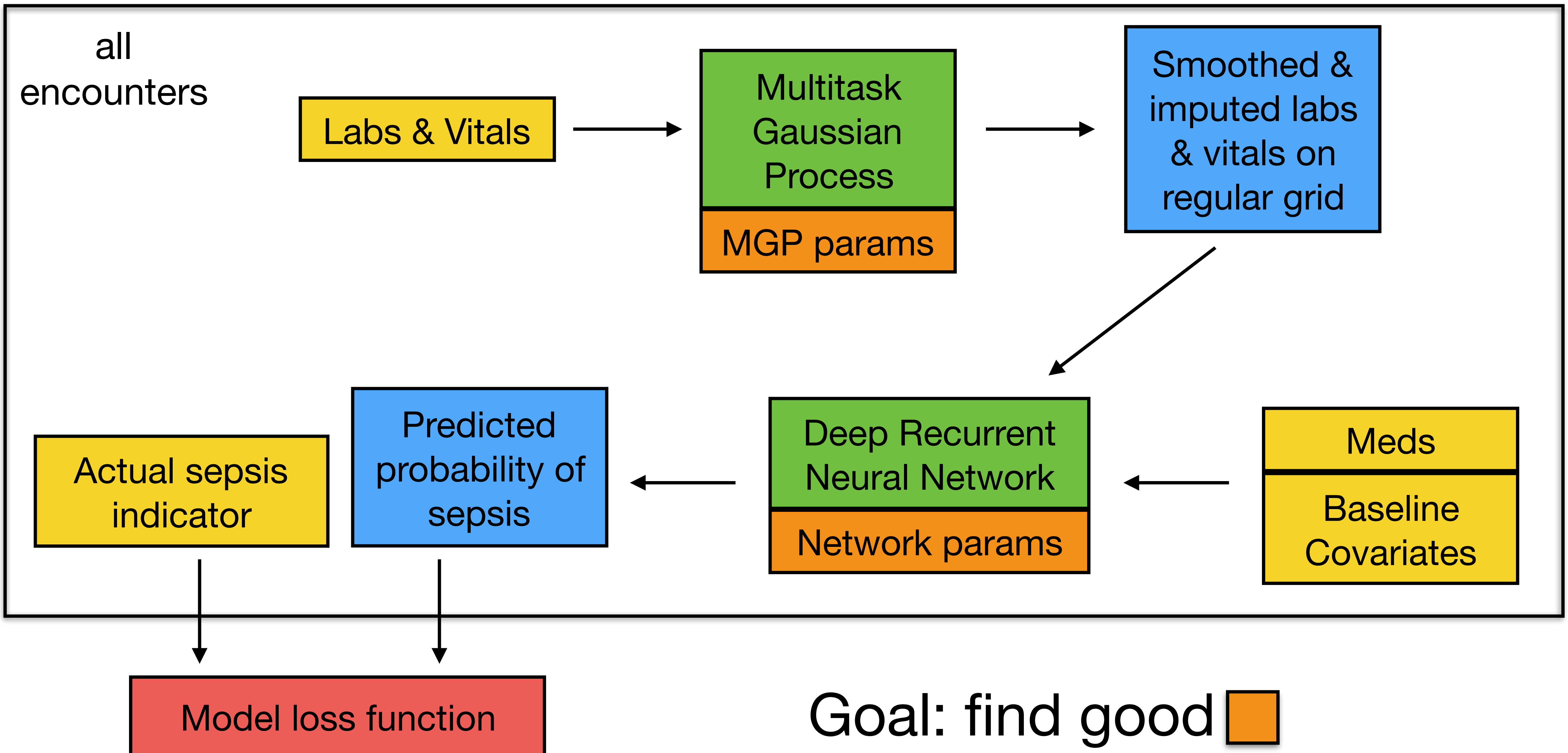
Early Warning Scores

National Early Warning Score (NEWS)*

PHYSIOLOGICAL PARAMETERS	3	2	1	0	1	2	3
Respiration Rate	≤8		9 - 11	12 - 20		21 - 24	≥25
Oxygen Saturations	≤91	92 - 93	94 - 95	≥96			
Any Supplemental Oxygen		Yes		No			
Temperature	≤35.0		35.1 - 36.0	36.1 - 38.0	38.1 - 39.0	≥39.1	
Systolic BP	≤90	91 - 100	101 - 110	111 - 219			≥220
Heart Rate	≤40		41 - 50	51 - 90	91 - 110	111 - 130	≥131
Level of Consciousness				A			V, P, or U

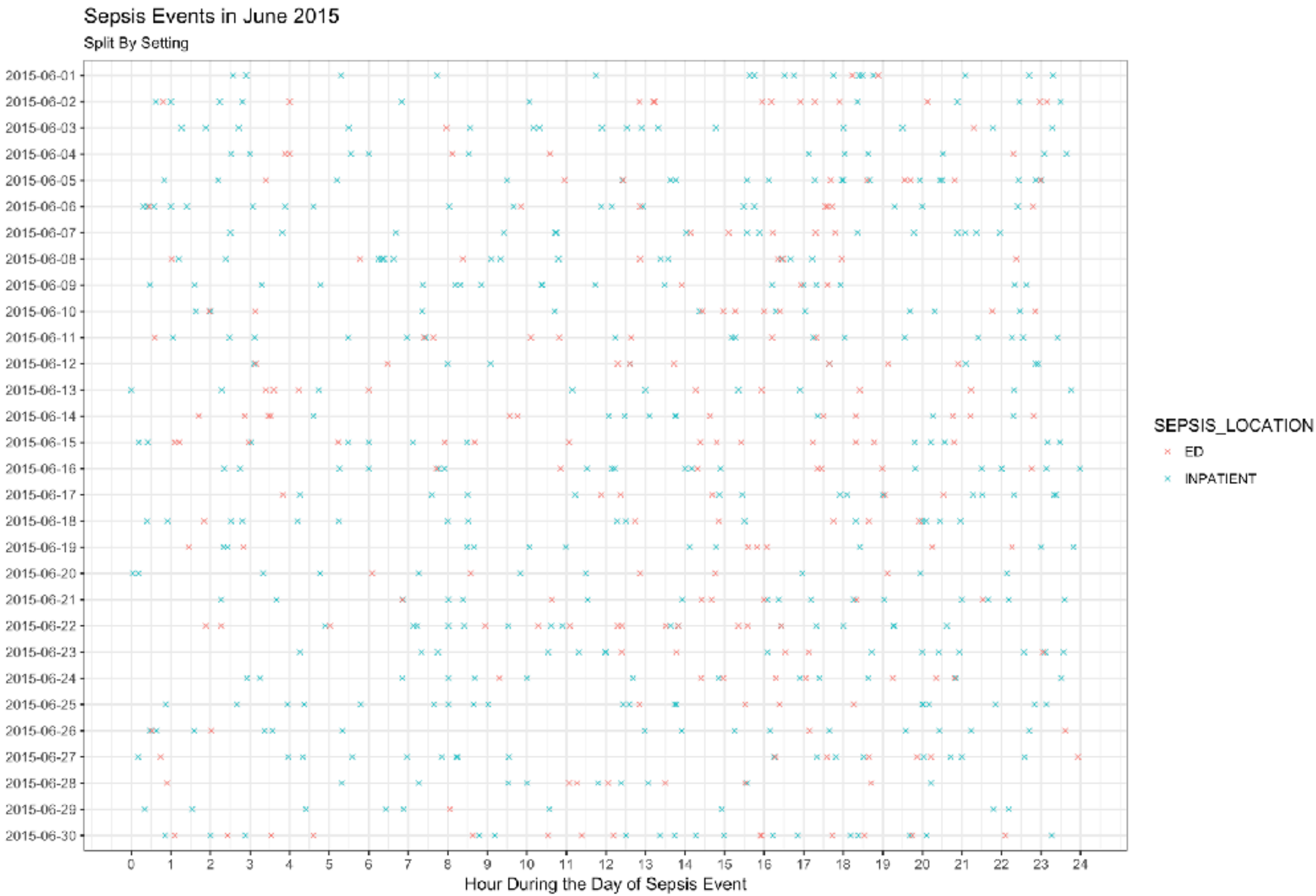
Can we do better?

Model Architecture



Goal: find good [orange box] to make [red box] small

End-to-end learning!



Sepsis Events in June 2015
Split By Setting
2015-06-01
2015-06-02
2015-06-03
2015-06-04
2015-06-05
2015-06-06
2015-06-07
2015-06-08
2015-06-09
2015-06-10
2015-06-11
2015-06-12
2015-06-13
2015-06-14
2015-06-15
2015-06-16
2015-06-17
2015-06-18
2015-06-19
2015-06-20
2015-06-21
2015-06-22
2015-06-23
2015-06-24
2015-06-25
2015-06-26
2015-06-27
2015-06-28
2015-06-29
2015-06-30
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Hour During the Day of Sepsis Event
SEPSIS_LOCATION
ED
INPATIENT

Results

- **MGP-RNN:** our approach
- **Raw RNN**: RNN trained on raw data (missing: carry forward last observed value)
- **PLR**: Penalized logistic regression, same imputation as Raw RNN
- **SIRS**, **NEWS**: clinical scores