Predictive Analytics for Tracking and Prognosis of Disease Activity

Suchi Saria
Assistant Professor
Computer Science, Applied Math & Stats and
Health Policy and Management
Institute for Computational Medicine

with **Kirill Dyagilev (CS), Peter Schulam (CS), Katie Henry (CS), David Hager (Critical Care), Peter Pronovost (Anesthesiology), Fredrick Wigley (Rheumatology), Daniel Robinson (Applied Math)**

Predictive Analytics

Example applications of predictive analytics

- Who is most likely to get re-admitted?
- Predict readiness for discharge
- Acute adverse event detection
- Prognosticating the course of disease in multi-morbid conditions

Example applications of predictive analytics

- Tailor therapies to the individual
- Early intervention
- Triage
- Resource Management (e.g., bed management, MRI scheduling, staff scheduling)
- ….

Theme: Methods for assessing the current and future health of an individual based on electronic health data.

Outline

Novel learning algorithms

- Current methods are highly susceptible to provider practice patterns.
- Interventional Confounds (ICs)
- New ranking based learning algorithm that avoids bias due to ICs.
- Application in the context of adverse event detection.

Extension to Practice-Cost Sensitive Learning

- Incorporate cost-of-practice into developing predictive models

Quick overview of other work

- Models for *personalized* prognosis in Complex, Chronic Conditions
- Smart phone based illness severity monitoring

Potentially Preventable Complications (PPC) lead to increased cost, length of stay, and mortality

PPCs add an estimated $88 billion to annual inpatient hospital costs nationally

Septicemia & severe infection

3.9 times higher mortality rate
2.4 times longer mean LOS
2.7 times higher mean cost

Acute lung edema and respiratory failure

5.2 times higher mortality rate
2.0 times longer mean LOS
2.3 times higher mean cost

Renal failure with dialysis

5.1 times higher mortality rate
2.5 times longer mean LOS
2.9 times higher mean cost

J. S. Hughes et al., Health Care Financing Review, vol. 27, pp. 63-82, 2006
R. L. Fuller, et al., Health Care Financing Review, vol. 30, pp. 17-32, 2009

Sepsis: Our Target Application

High mortality and morbidity

Septicemia is the 11th leading cause of death in the US

- Over 750,000 cases of severe sepsis or septic shock annually and rising
- Mortality of septic shock estimated between 30-60%
- Patients with sepsis have increased hospital stays and long-term morbidities

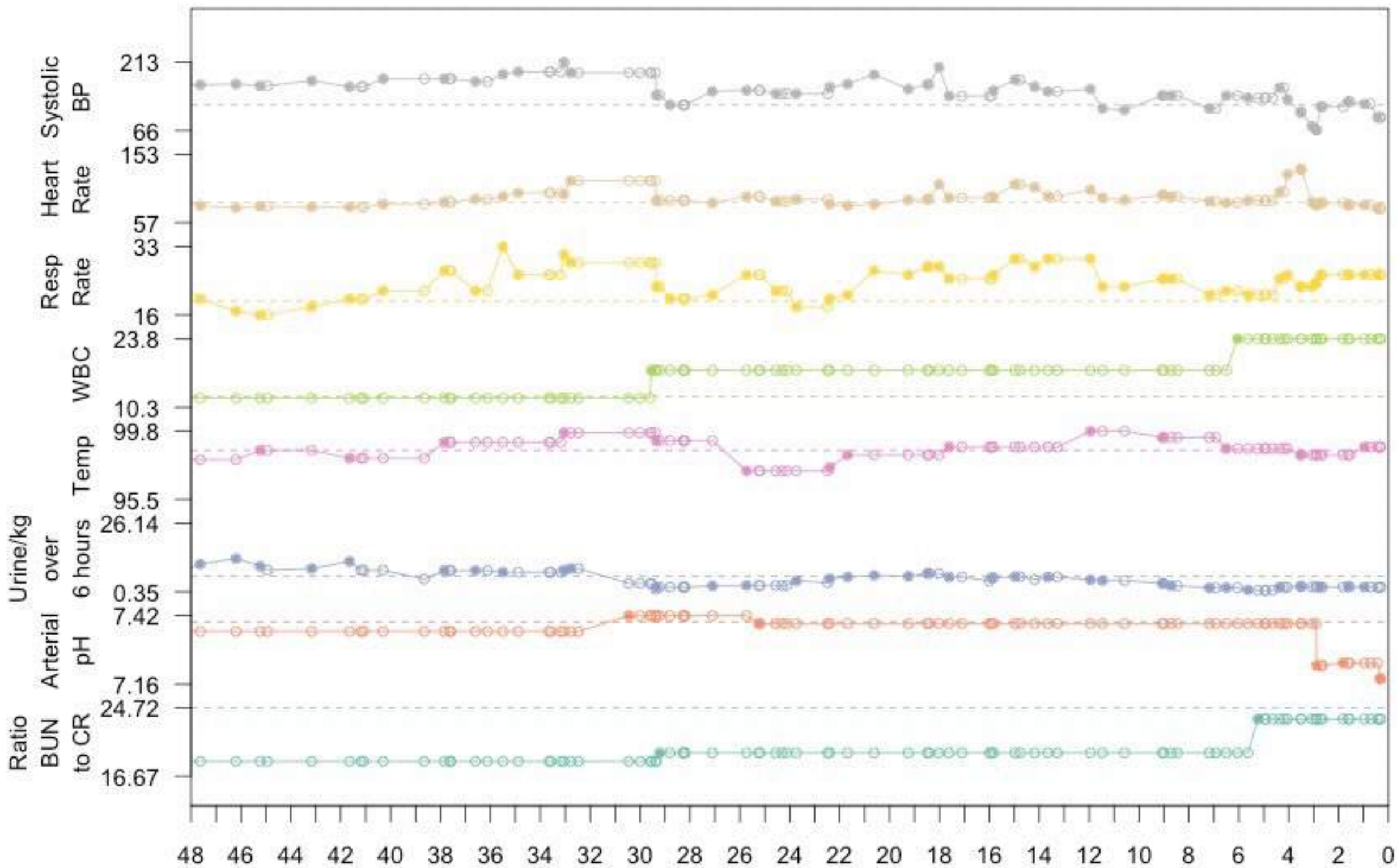
High financial cost

- $15.4 billion dollars spent nationally

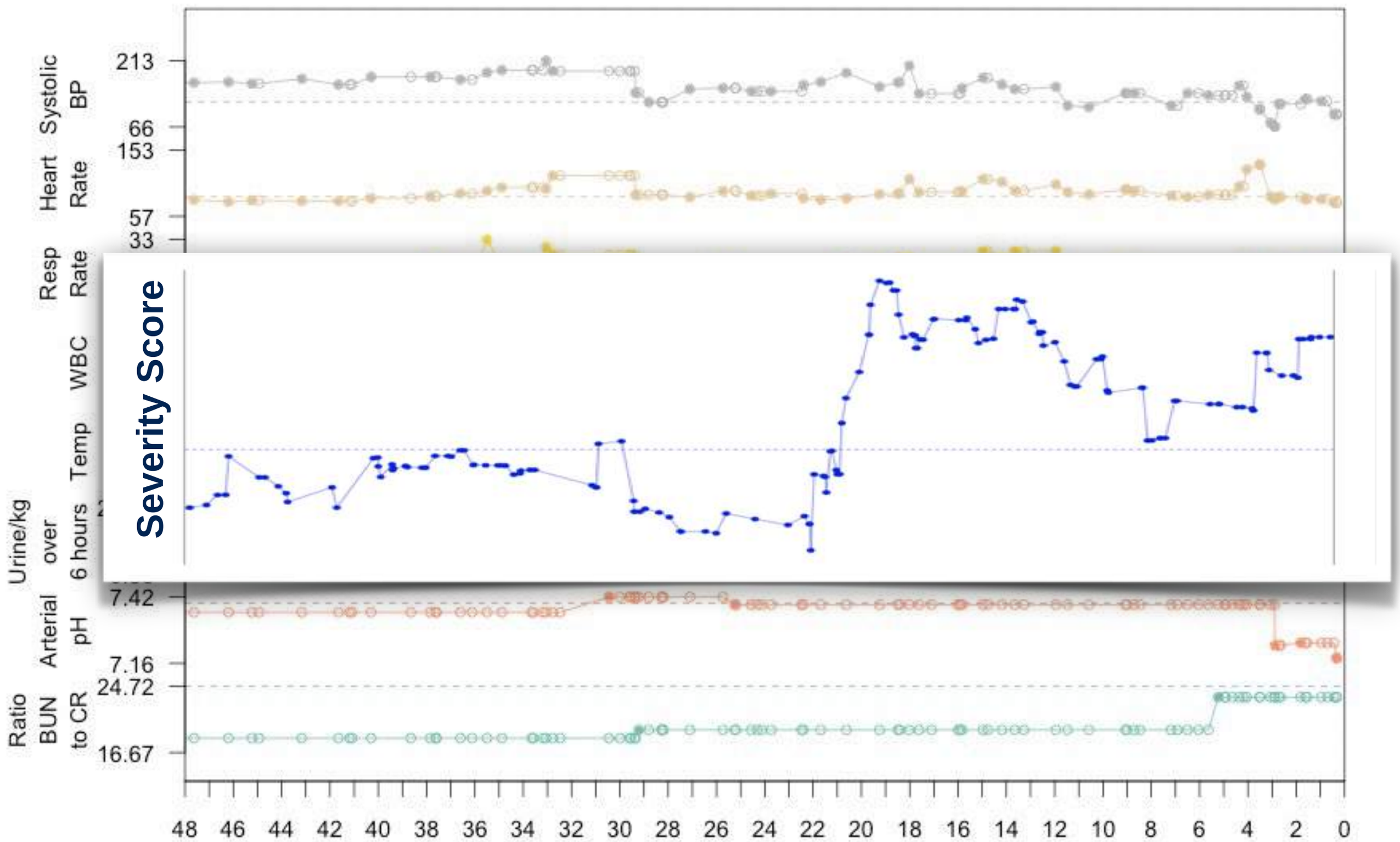
Mortality and length of stay decreased with timely treatment [Kumar et al. 2011]

P. R. Dellinger et al. Critical Care Medicine, vol. 41, no. 2, pp. 580–637, 2013
D. C. Angus et al., Critical Care Medicine, vol. 29, no. 7, 2001.
G. Kumar et al., Chest, vol. 140, no. 5, pp. 1223–1231, 2011.
HCUP Facts and Figures: Statistics on Hospital-based Care in the United States, 2009.

Scalable Continuous Monitoring



Scalable Continuous Monitoring



APACHE: measuring in-hospital severity

[7 bins]

...

110 < Heart rate < 139 bpm	+2 points
140 < Heart rate > 179 bpm	+3 points
Heart rate > 180 bpm	+4 points

[5 bins]

...

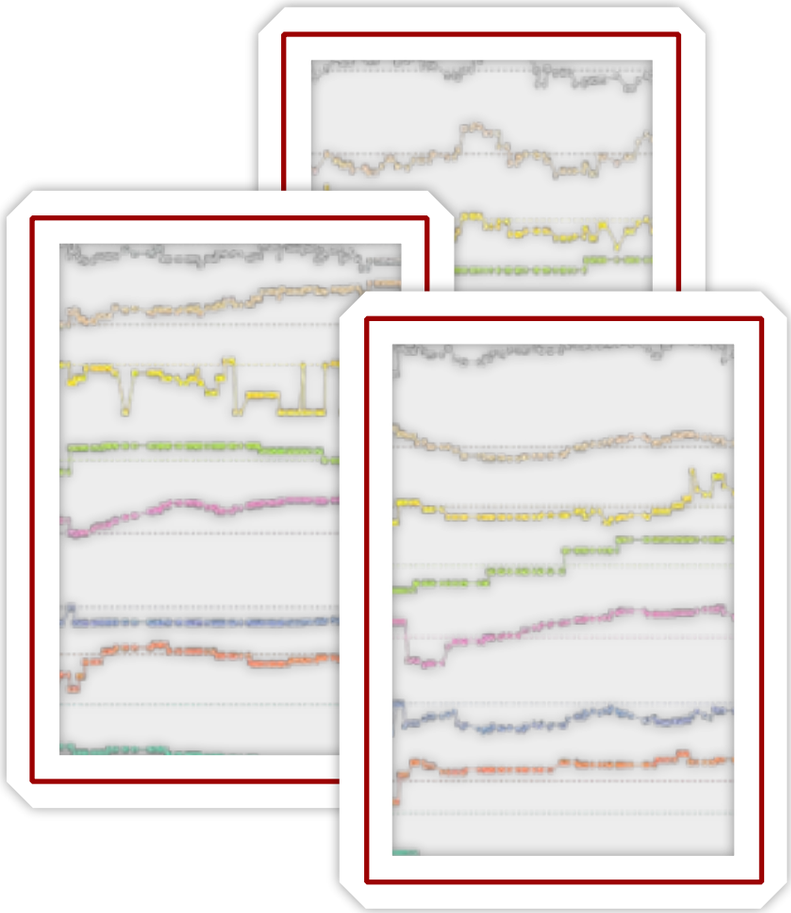
55 < Age < 64 bpm	+3 points
65 < Age < 74 bpm	+5 points
Age >=75 bpm	+6 points

Severity is a measure of total damage of the disease on the body [Medsger, 2003]
Also interpreted as the counterfactual risk of decline or adverse events (e.g., mortality)

Medsger et al., Clinical and Experimental Rheumatology, 2003
Knaus et al, Critical Care Medicine, Vol. 13, Issue 10, 1985

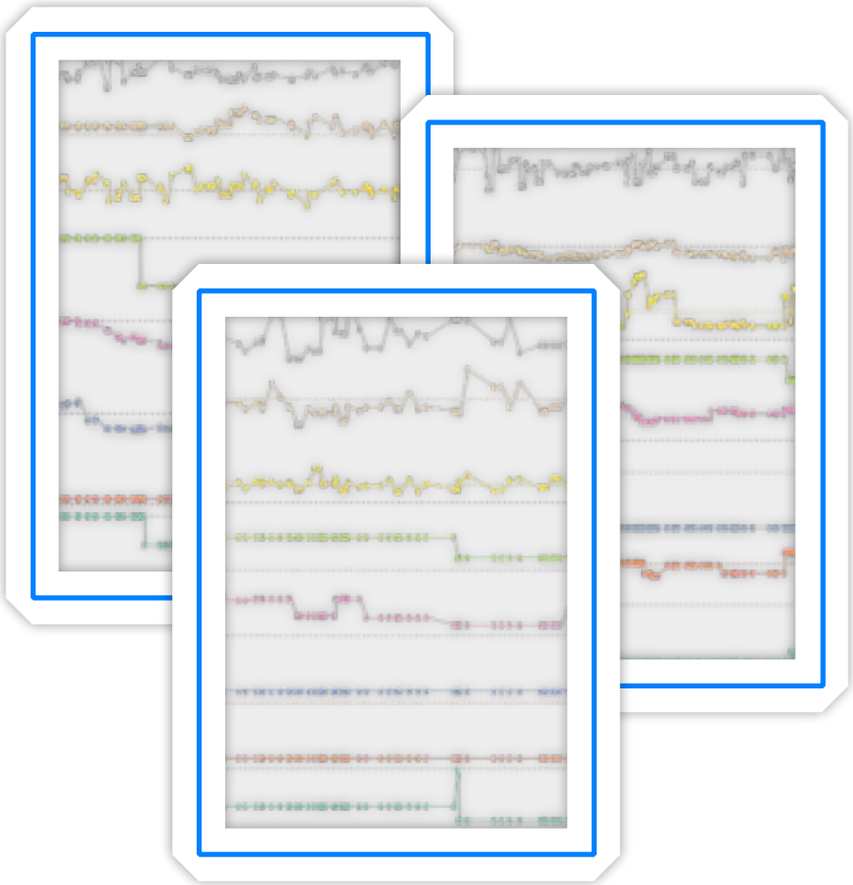
Predictive Model for Downstream Adverse Event

Automatically learn patterns that distinguish patients with AE from those without



patients with Adverse Event

VS



patients without Adverse Event

Pneumonia Severity Index: Risk of Mortality

- Identify candidate risk factors
- Learn score and relative weights by regressing against observed mortality

Demographics	Co-morbidities	Physical exam / vital signs	Laboratory / imaging
▪ Age (1 point per year) Male Yr Female Yr -10 ▪ Nursing home residency +10	▪ Neoplasia +30 ▪ Liver disease +20 ▪ CHF +10 ▪ Cerebrovascular disease +10 ▪ Renal disease +10	▪ Mental confusion +20 ▪ Respiratory rate +20 ▪ SBP +20 ▪ Temperature +15 ▪ Tachycardia +15	▪ Arterial pH +30 ▪ BUN +20 ▪ Sodium +20 ▪ Glucose +10 ▪ Hematocrit +10 ▪ Pleural effusion +10 ▪ Oxygenation +10

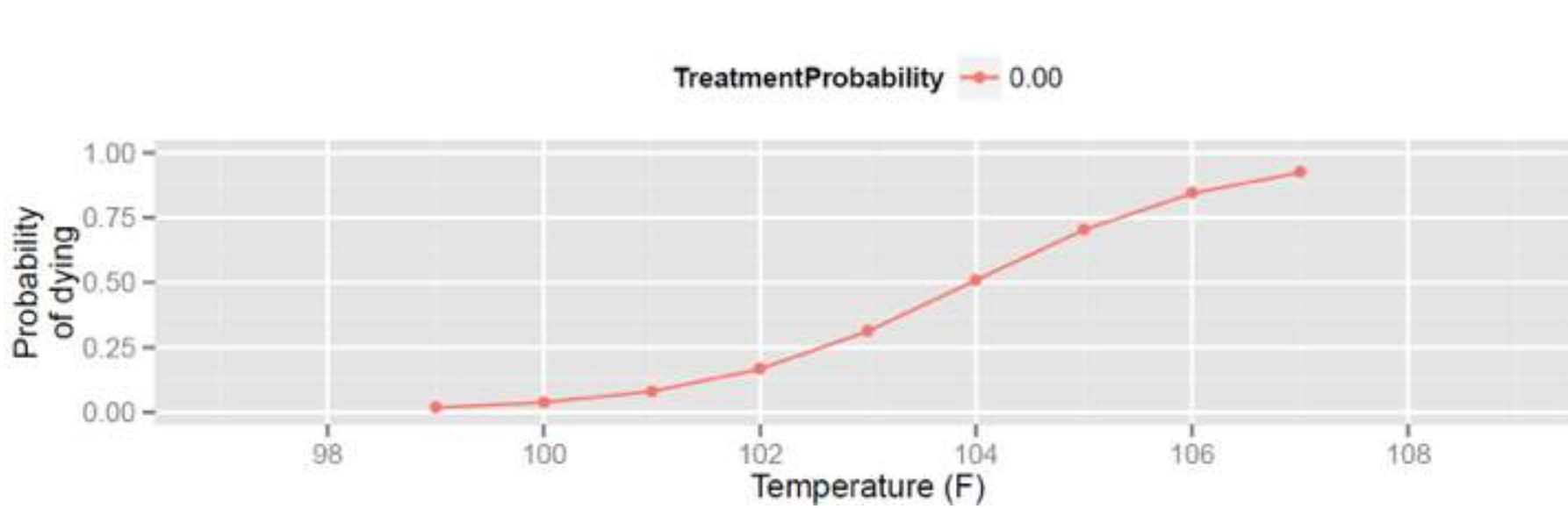
Risk class (Points)	Mortality (%)	Recommended site of care
I (<50)	0.1	Outpatient
II (51–70)	0.6	Outpatient
III (71–90)	2.8	Outpatient or brief inpatient
IV (91–130)	8.2	Inpatient
V (>130)	29.2	Inpatient

Fine et al., *N Engl J Med* 336 (4): 243–250, 1997

Learns Risk Estimates Sensitive to Provider Practice Pattern

- Simple example (Flu)
  - Measure temperature
  - Measure WBC

- Increase in temperature or WBC increases risk of death



Learns Risk Estimates Sensitive to Provider Practice Pattern

Key idea #1:

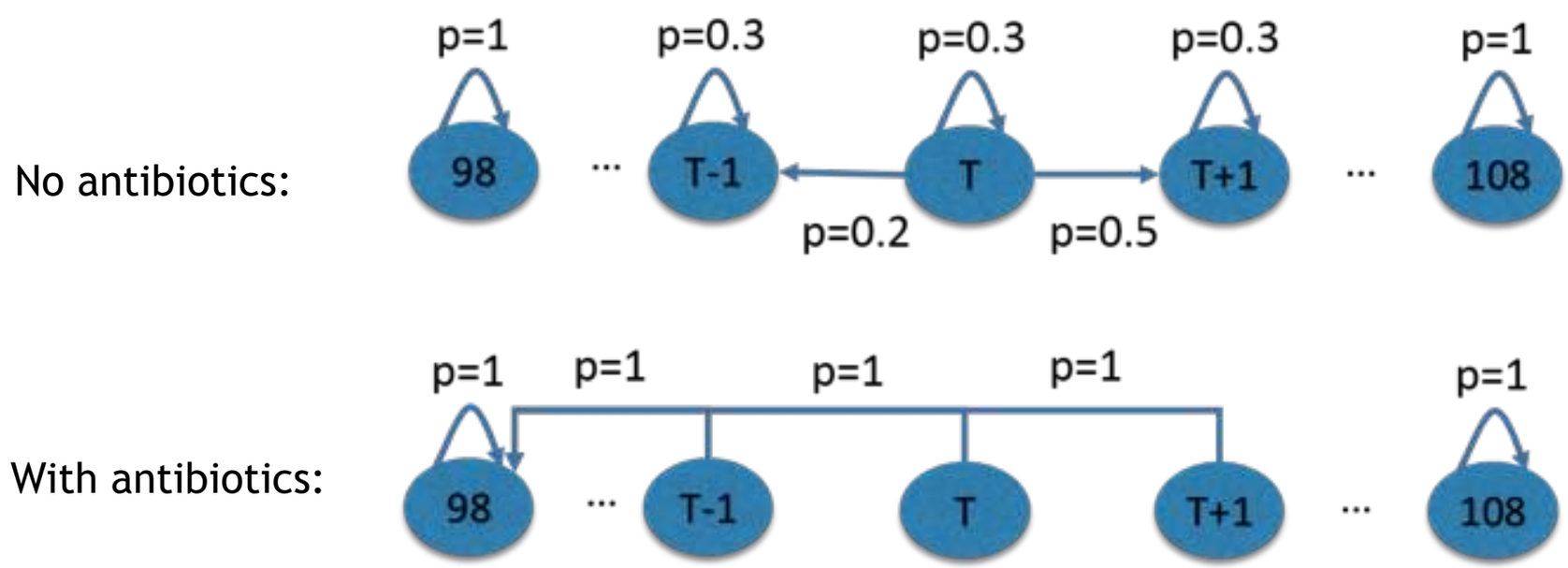
- Consider a unit where patients get treated as temperature increases above say, 102 degrees
  - Therefore, *fewer deaths due to rising temperature*
  - As fewer individuals experience death, the **algorithm no longer associates rise in temperature with risk.**

Bias due to interventional Confounds

C. Paxton, A. Niculescu-Mizil, S. Saria. **Developing Predictive Algorithms Using Electronic Medical Records: Challenges and Pitfalls**. American Medical Informatics Association, 2013. pdf

Bias Due to Interventional Confounds

- Model flu severity; temperature is observed
- Example: Synthetic-Pneumonia
  - If flu, temperature increases unless medicated
  - When medicated, temperature returns to normal
  - At 108 deg F, subject dies
- Consider hospitals with different practice patterns: P(med | temperature)



Treatment practice:
(1) no antibiotics for $T < 102$ deg F;
(2) administer antibiotics with probability ρ for $T \geq 102$ deg F

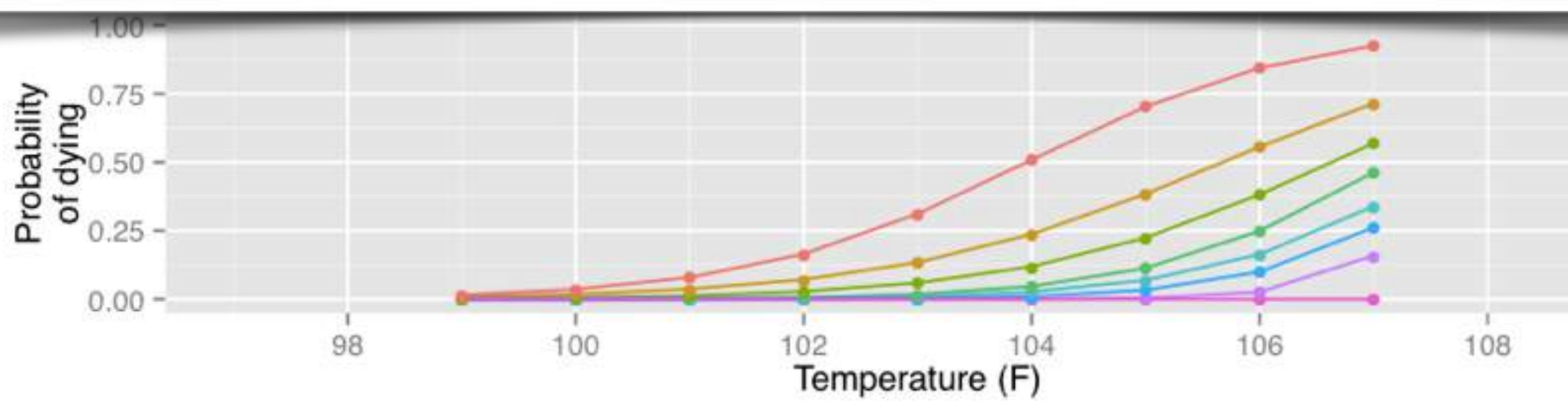
Medsger et al., Clinical and Experimental Rheumatology, 2003

Bias Due to Interventional Confounds

- Model flu severity; temperature is observed

- Example: Synthetic-Pneumonia
  - If flu, temperature increases unless medicated
  - When medicated, temperature returns to normal

- Learned risk score is sensitive to practice patterns
- 103 deg F may be considered severe or benign based on practice patterns => violates construct validity [Medsger, 2003]
- May worsen outcomes if used as decision aid

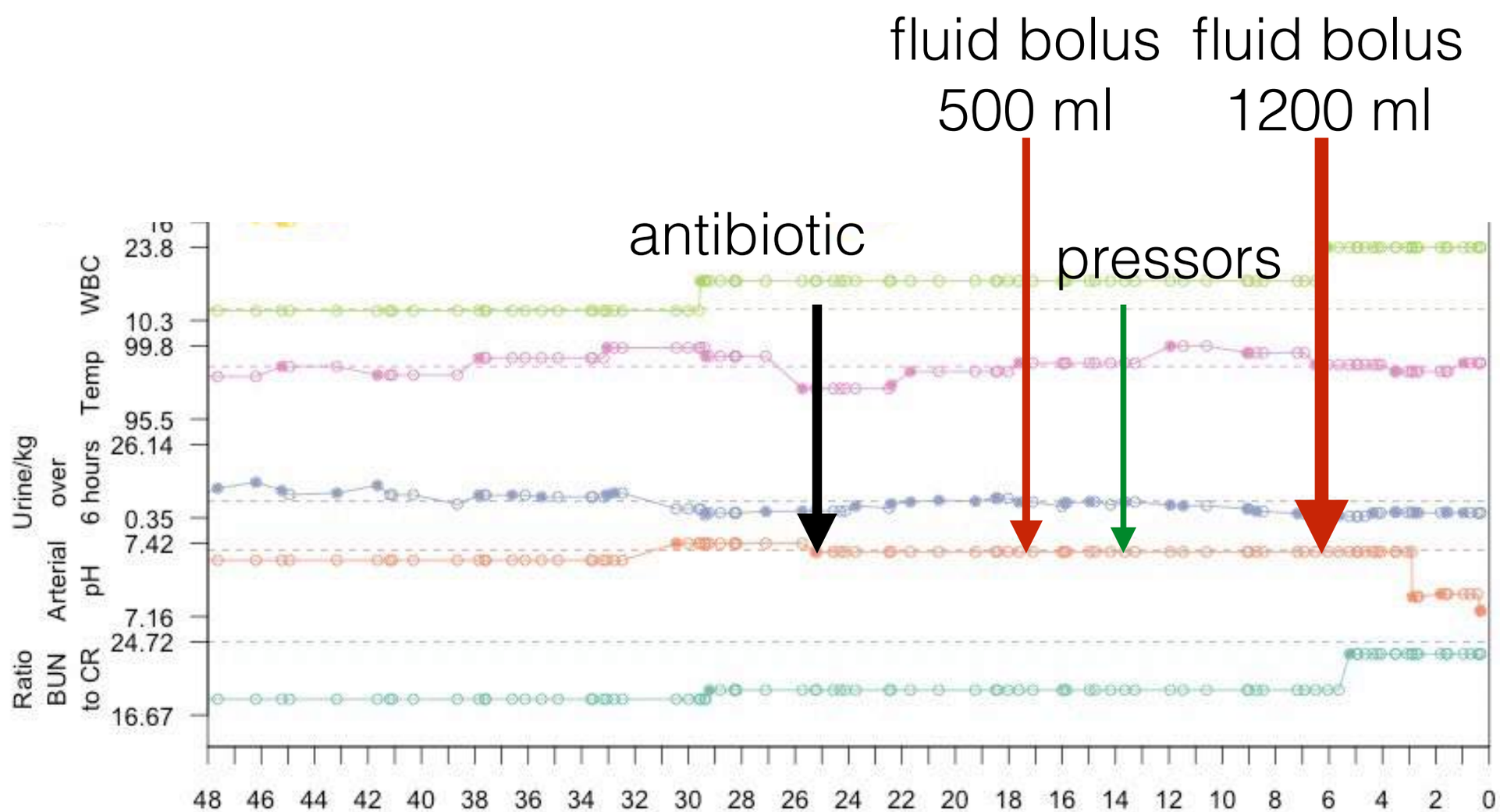


See e.g. for bias due to interventional confounds in **Paxton et al., American Medical Informatics Association, 2011**

Medsger et al., Clinical and Experimental Rheumatology, 2003

Probabilistic State Estimation Approaches

Treat the severity state as latent and develop a causal model estimate state given the observed measurements.



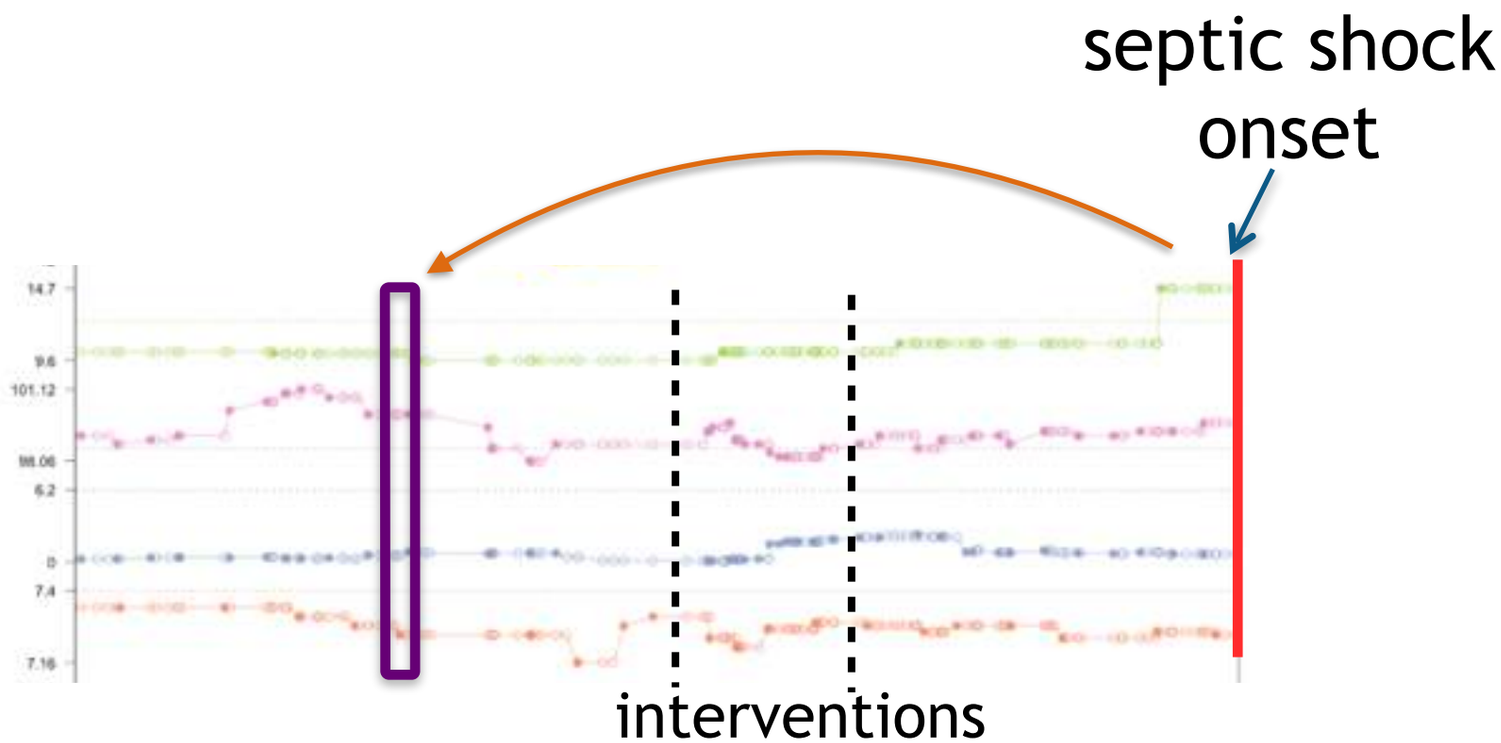
Robins et al., Epidemiology, Vol. 11, No. 5, 2000

Mould. Clinical Pharmacology & Therapeutics 92(1):125–131. 2012

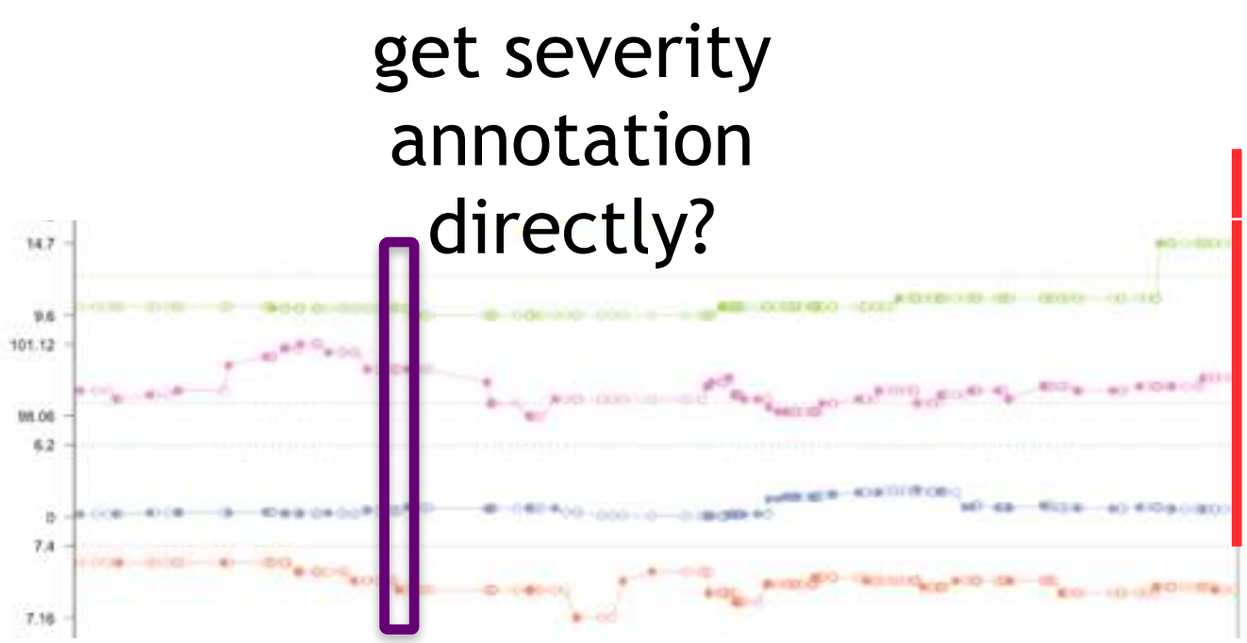
S Saria et al., 2010. **Integration of Early Physiological Responses Predicts Later Illness Severity in Preterm Infants**. Science *Translational Medicine*, September 2010. Vol. 2, Issue 48

S. Saria et al., 2010. **Learning individual and population level traits from clinical temporal data.** In *Neural Information Processing Systems (NIPS), Predictive Models in Personalized Medicine workshop*. 2010.

Alternate source of supervision?



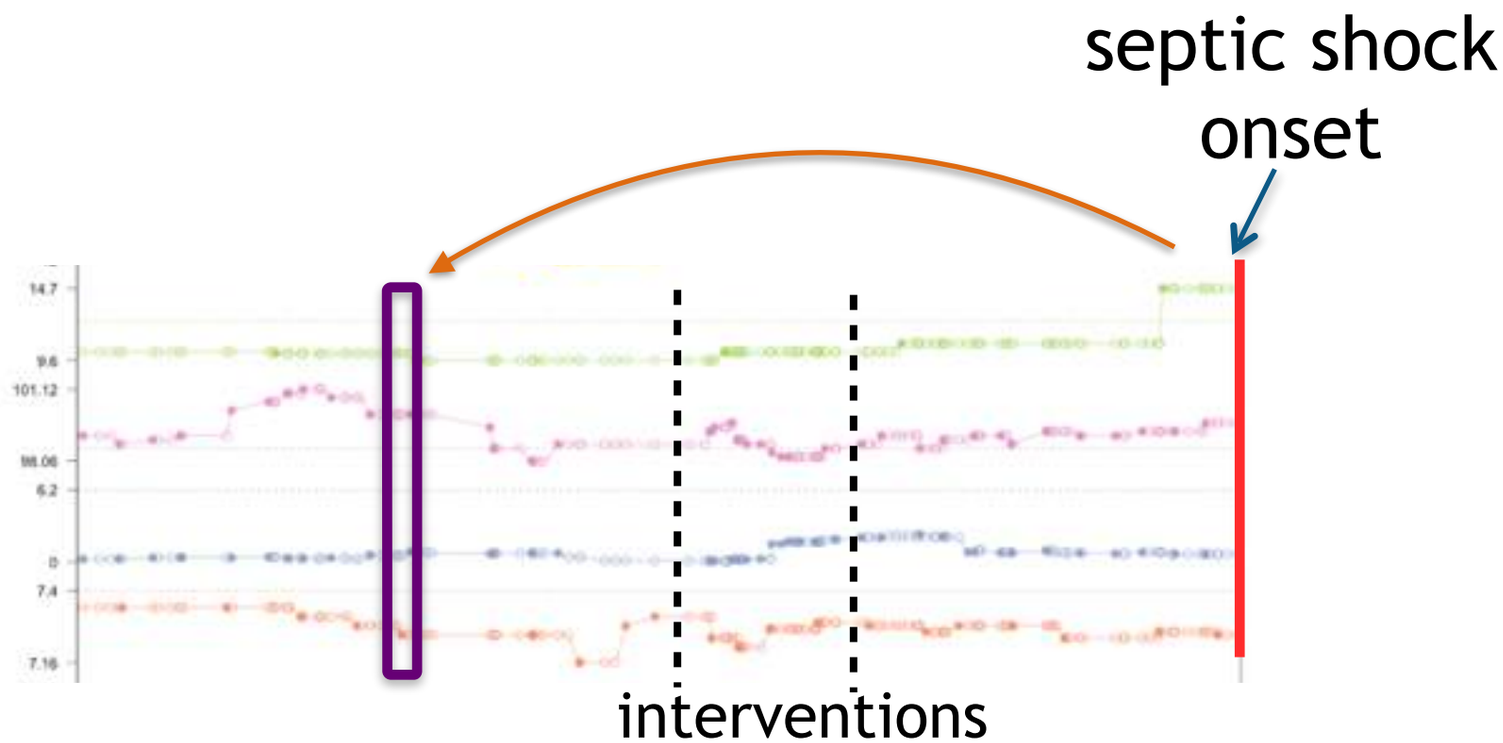
Instead:



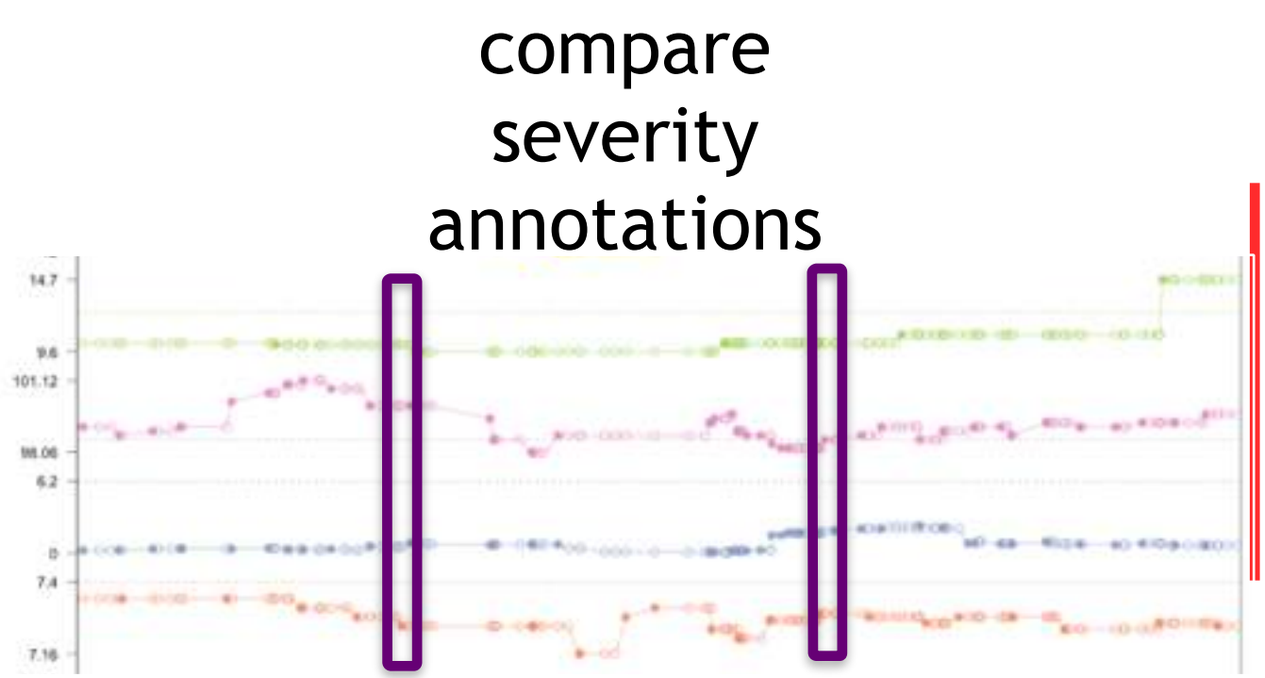
Medsger et al., Clinical and Experimental Rheumatology, 2003

P. R. Dellinger et al. Critical Care Medicine, vol. 41, no. 2, pp. 580–637, 2013

Alternate source of supervision?



Instead:



Medsger et al., Clinical and Experimental Rheumatology, 2003

P. R. Dellinger et al. Critical Care Medicine, vol. 41, no. 2, pp. 580–637, 2013

Sepsis: Our Target Application

SIRS
- Systemic Inflammatory Response Syndrome

→

Sepsis
- SIRS with suspected infection

→

Severe Sepsis
- Sepsis with organ dysfunction

→

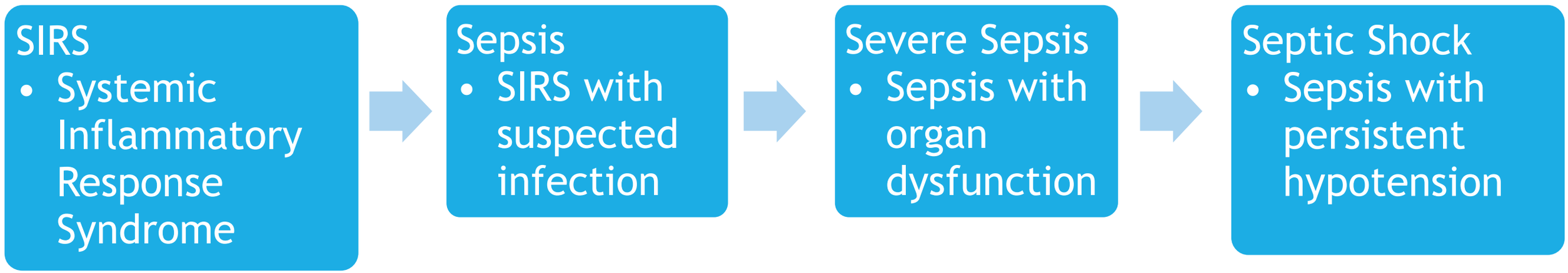
Septic Shock
- Sepsis with persistent hypotension

***Clinical Comparisons*:**

- Comparing the ***counterfactual risk of decline*** for two patient states
- If left untreated, risk of immediate decline in subject w/ septic shock > risk of immediate decline in subject w/ SIRS

Medsger et al., Clinical and Experimental Rheumatology, 2003
P. R. Dellinger et al. Critical Care Medicine, vol. 41, no. 2, pp. 580–637, 2013

Sepsis: Our Target Application



Guideline

- Coarse severity levels of sepsis
- Recognize instances of each stage using data in the electronic health record
  - Combine natural language process of text notes with criteria using vital signs and laboratory test results
  - e.g., Septic Shock: If suspicion of infection AND two of the following (anomolous temp, RR, HR, WBC) AND one of the following (BP > 90 Hgmm despite fluids > 500 ml OR lactates >= 4)

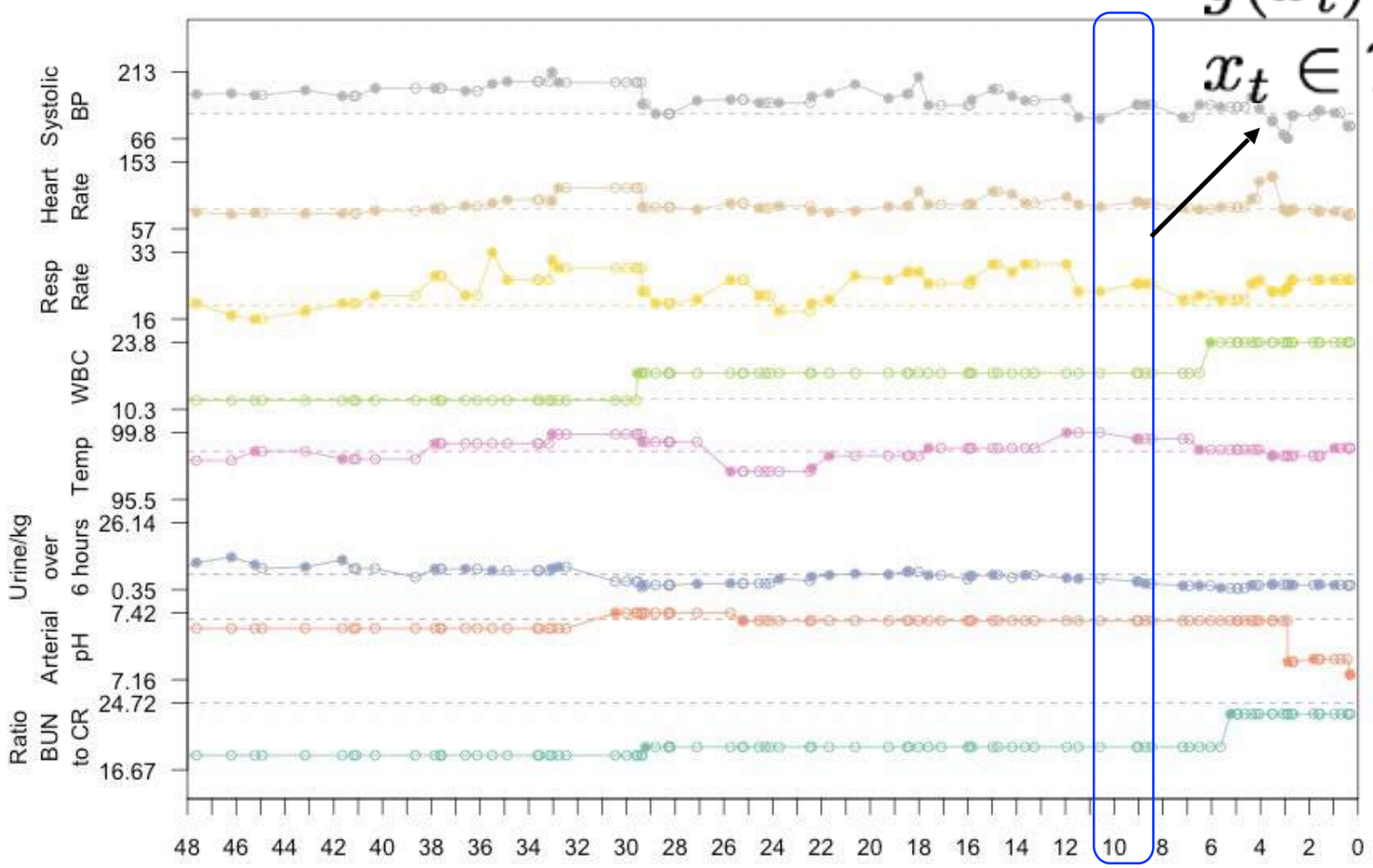
P. R. Dellinger et al. Critical Care Medicine, vol. 41, no. 2, pp. 580–637, 2013

Learning from Clinical Comparison Pairs

- Learn a function that maps measurements to severity s.t.
  - Concordant with experts ranking of severity
  - The underlying score is temporally smooth —> satisfies clinical expectations of how quickly the disease evolves.

$$g(x_t) = s_t$$

$$x_t \in \mathcal{R}^d \text{ and } s_t \in \mathcal{R}$$

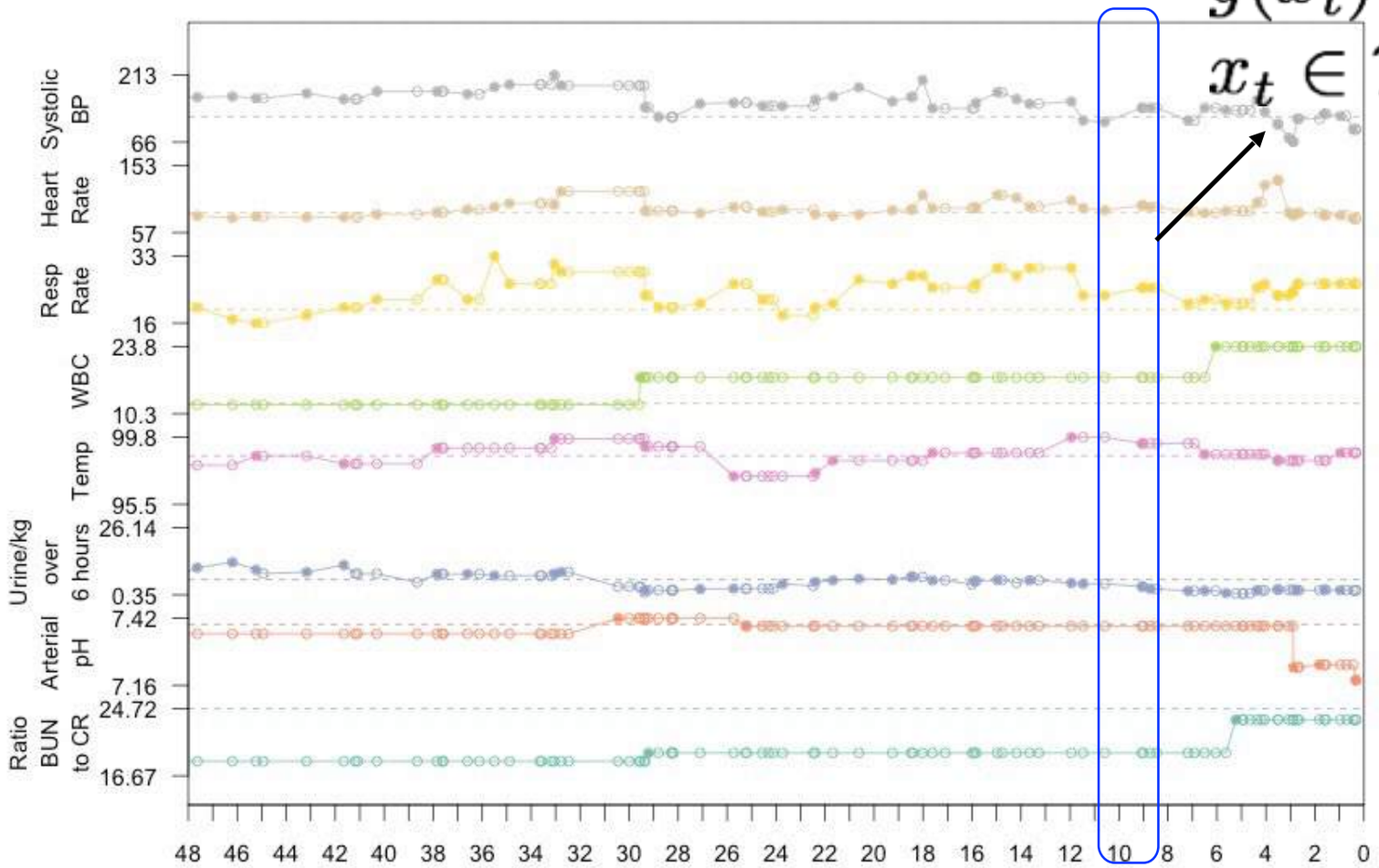


Learning from Clinical Comparison Pairs

- Max-margin formulation
  1. Maximize margin between pairs

$$\forall(< x_i^p, t_i^p >, < x_j^q, t_j^q >) \in O: \quad g_w(x_i^p) - g_w(x_j^q) \geq \mu^{(p,i),(q,j)}$$
$$\mu^{(p,i),(q,j)} > 0$$

$$g(x_t) = s_t$$
$$x_t \in \mathcal{R}^d \text{ and } s_t \in \mathcal{R}$$



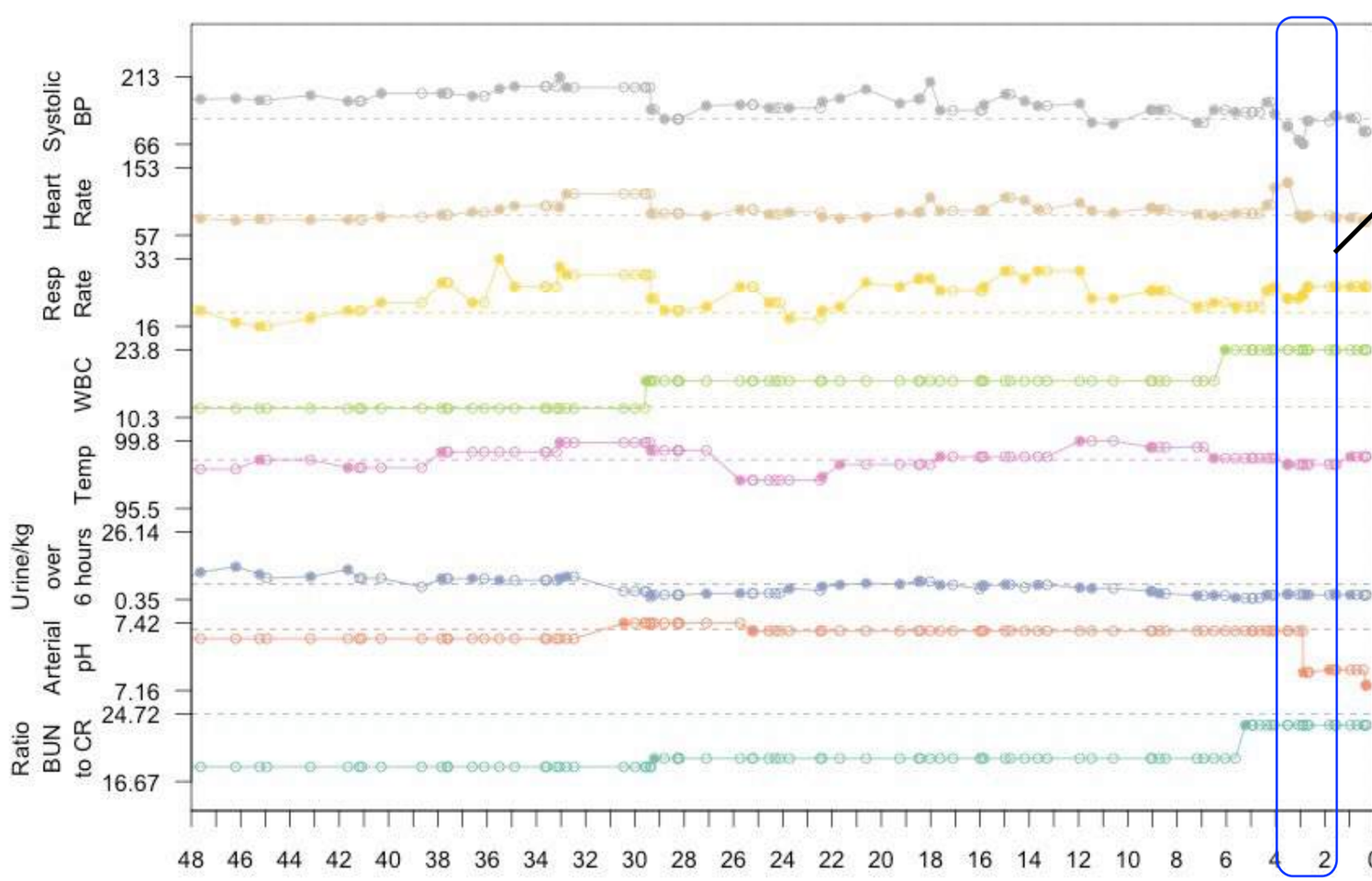
Learning from Clinical Comparison Pairs

- Max-margin formulation
  1. Maximize margin between pairs

$$\forall(< x_i^p, t_i^p >, < x_j^q, t_j^q >) \in O: \quad g_w(x_i^p) - g_w(x_j^q) \geq \mu^{(p,i),(q,j)}$$
$$\mu^{(p,i),(q,j)} > 0$$

  2. Smooth trajectories

$$\sum_{(<x_i^p,t_i^p>,<x_{i+1}^p,t_{i+1}^p>)\in S} \left[\frac{g(x_{i+1}^p)-g(x_i^p)}{t_{i+1}^p-t_i^p}\right]^2$$



$$g(x_t) = s_t$$
$$x_t \in \mathcal{R}^d \text{ and } s_t \in \mathcal{R}$$

L-DSS: Linear Disease Severity Score

- Convex, twice-differentiable.
- Optimized using the Newton-Raphson algorithm

L-DSS Objective:

$$\min_{w} \quad \frac{1}{2}\|w\|^2 + \frac{\lambda_{\text{ord}}}{|O|} \sum_{(<\mathbf{x}_i^p, t_i^p>, <\mathbf{x}_j^q, t_j^q>) \in O} L_h(1 - w^T(\mathbf{x}_i^p - \mathbf{x}_j^q)) + \frac{\lambda_{\text{smooth}}}{|S|} \sum_{(<\mathbf{x}_i^p, t_i^p>, <\mathbf{x}_{i+1}^p, t_{i+1}^p>) \in S} \left[\frac{w^T(\mathbf{x}_{i+1}^p - \mathbf{x}_i^p)}{t_{i+1}^p - t_i^p}\right]^2$$

Joachims. Knowledge Discovery and Data Mining, 2003
Chapelle and Keerthi. Information Retrieval, 2010
Chapelle. Neural Computation, 2007

Dataset and Population

MIMIC II Clinical Database

- **Electronic health records (EHR) collected from intensive care units as the Beth Israel Deaconess Medical Center from 2001 to 2008**

Study population

- **16,232 adult patients included (age ≥ 15 years)**
- **60% Training set**
- **40% Test set**

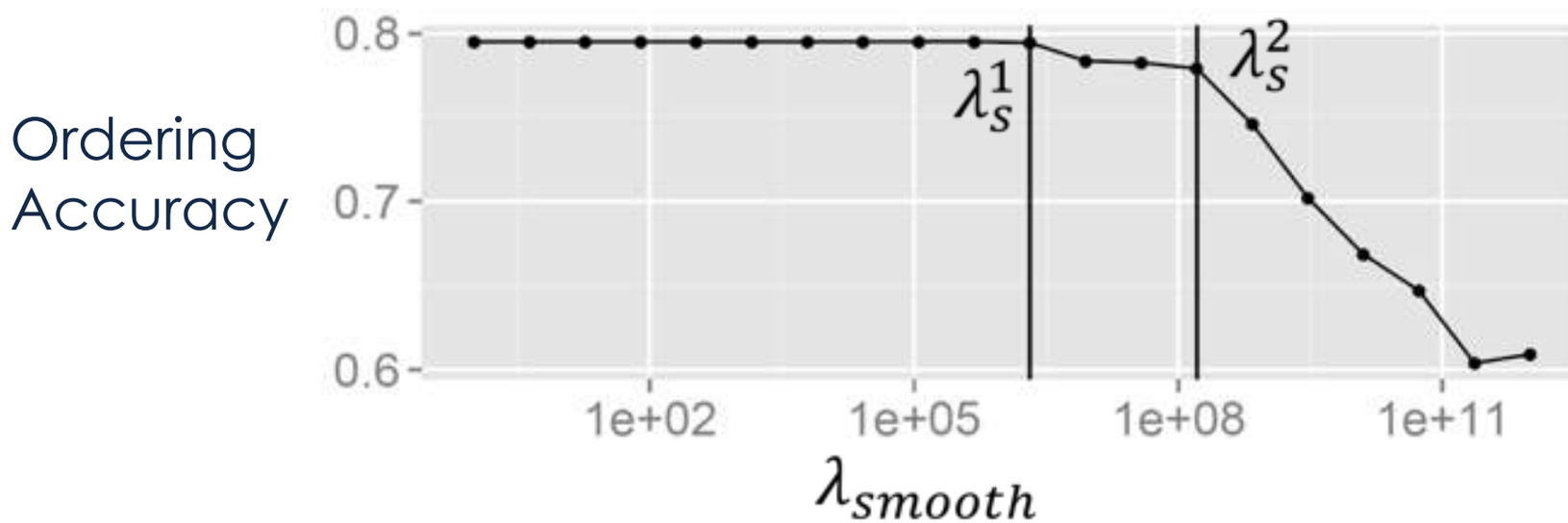
A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic sig- nals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13).
R. Dellinger *et al.* "Surviving sepsis campaign: International guidelines for management of severe sepsis and septic shock," *Critical Care Medicine*, vol. 41, no. 2, pp. 580–637, 2013.

Selection of λ_{ord} and λ_{smooth}

L-DSS Objective :

$$\min_{w} \quad \frac{1}{2}\|w\|^2 + \frac{\lambda_{\text{ord}}}{|O|} \sum_{(<\mathbf{x}_i^p, t_i^p>, <\mathbf{x}_j^q, t_j^q>) \in O} L_h(1 - w^T(\mathbf{x}_i^p - \mathbf{x}_j^q)) + \frac{\lambda_{\text{smooth}}}{|S|} \sum_{(<\mathbf{x}_i^p, t_i^p>, <\mathbf{x}_{i+1}^p, t_{i+1}^p>) \in S} \left[\frac{w^T(\mathbf{x}_{i+1}^p - \mathbf{x}_i^p)}{t_{i+1}^p - t_i^p}\right]^2$$

Identify λ_{ord} by setting λ_{smooth} to zero.



Choice of λ_{smooth}

- λ_s^1 - maximal smoothness with the same ordering accuracy as λ_{smooth} = 0.
- λ_s^2 - optimizer of early detection performance (will be elaborated on later).

Numerical Analyses

1. **Ranking performance: does the proposed score order for severity more accurately than existing scores?**

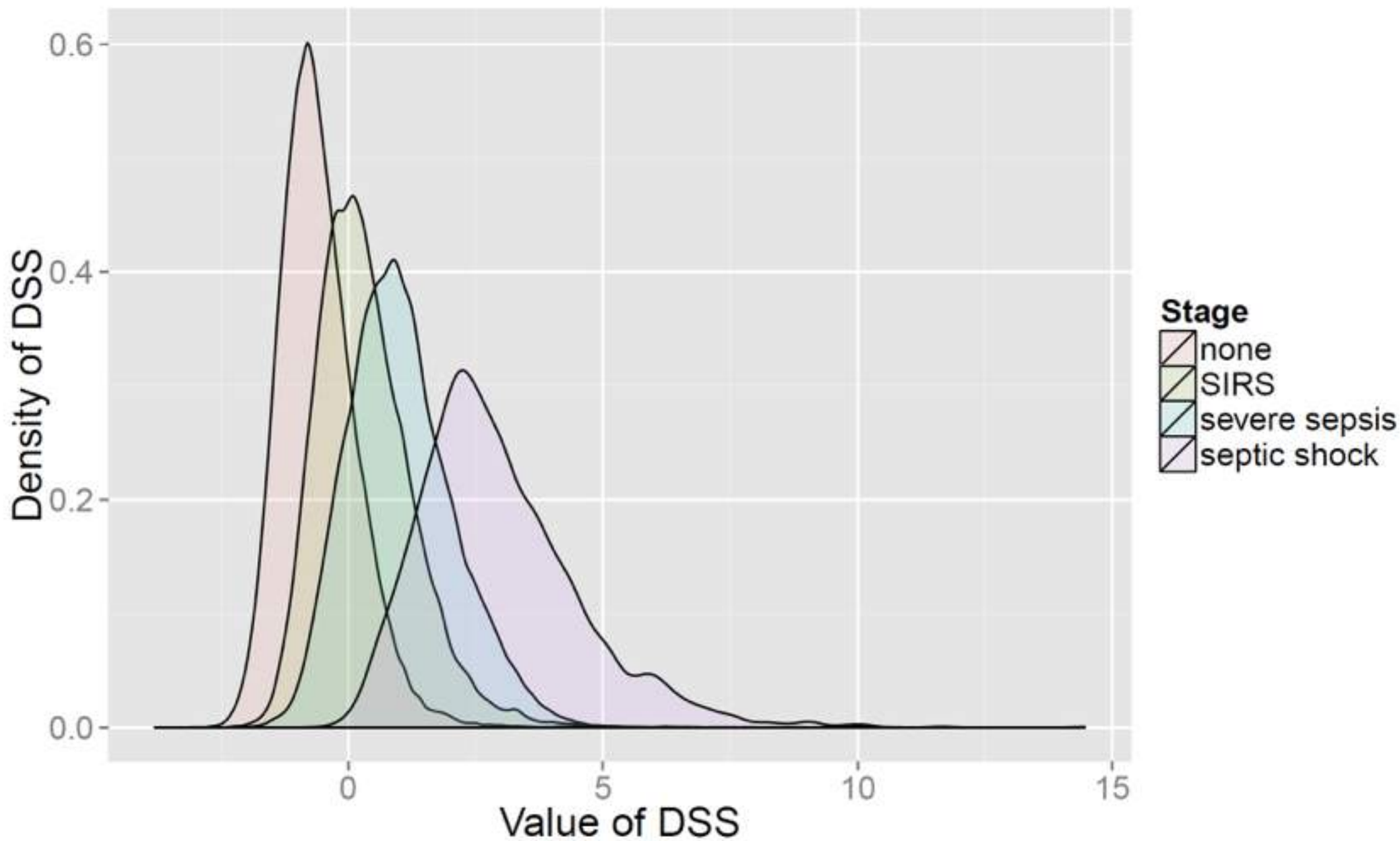
2. **How sensitive is the score to more granular changes in severity beyond coarse grading for the stages of sepsis?**

1. **Sensitive to impending adverse events: Leading unto an adverse event (e.g., septic shock), does the score reflect increase in severity?**

   - **Predictive performance: Measure prediction performance of the resulting score.**

2. **Sensitive to therapy: Following a therapy (e.g., fluid bolus), does the score reflect decrease in severity?**

Learned DSS for $\lambda_{smooth} = \lambda_s^1$



E.g., Septic Shock among DSS > 2 is 20 times more frequent than in the "general population" of all data points

Task 1: Does DSS rank patient states with known severity ordering correctly?

- Metric: fraction of correctly ordered pairs on held out data
- Establish construct validity (Medsger 2003): consistent with clinical expectations

Type	Predictor	λ_{smooth}	Fraction of correctly ordered pairs
Proposed Scores	*L-DSS*	λ_s^1	***0.88±0.01***
	L-DSS	λ_s^2	***0.84±0.01***
Routine *Clinical Scores*	*APACHE II*	-	*0.68±0.01*
	Worst SOFA	-	*0.63±0.01*
	Total SOFA	-	*0.63±0.01*

Numerical Analyses

1. **Ranking performance: does the proposed score order for severity more accurately than existing scores?**

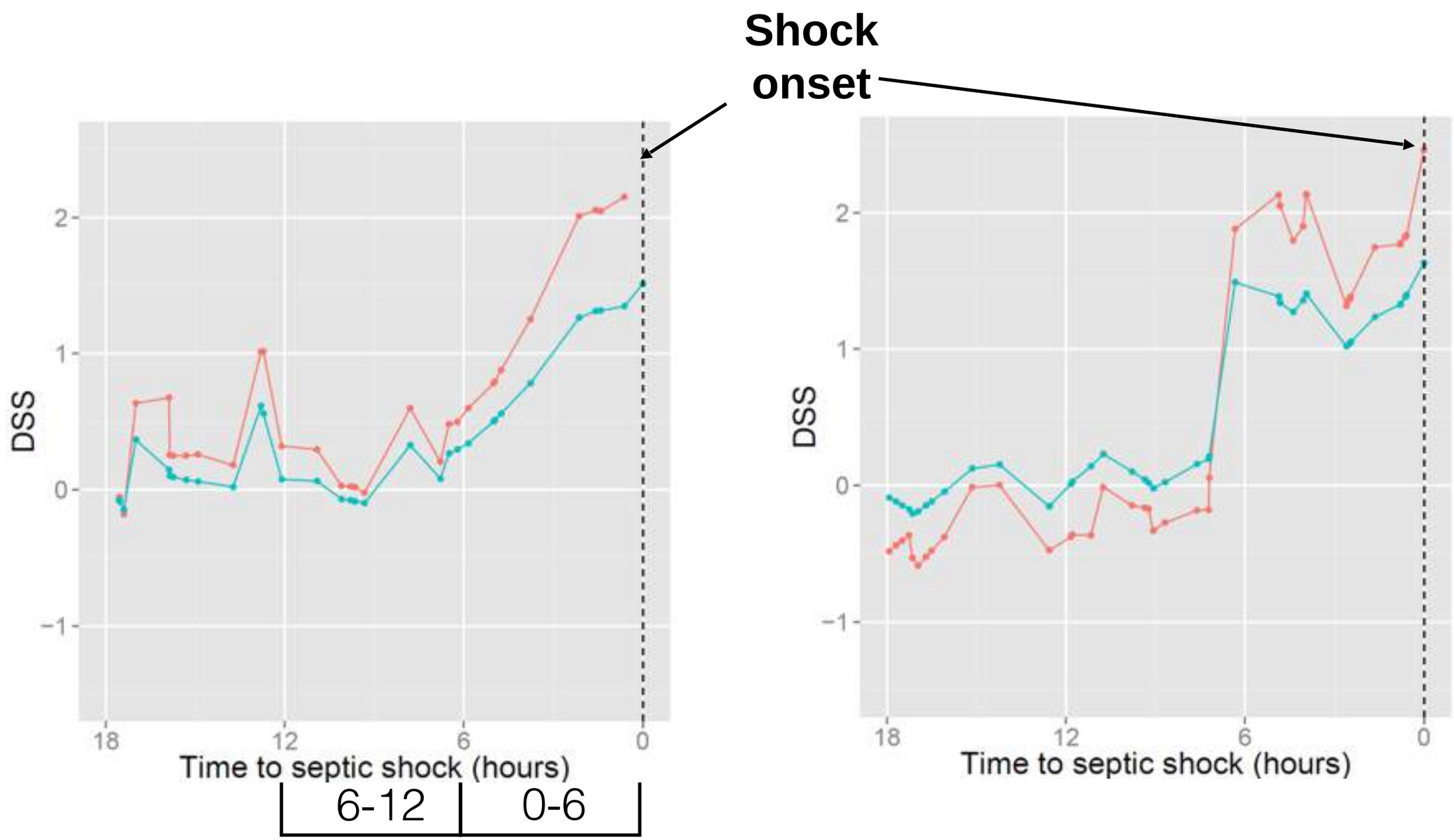
2. **How sensitive is the score to more granular changes in severity beyond coarse grading for the stages of sepsis?**

1. **Sensitive to impending adverse events: Leading unto an adverse event (e.g., septic shock), does the score reflect increase in severity?**

- **Predictive performance: Measure prediction performance of the resulting score.**

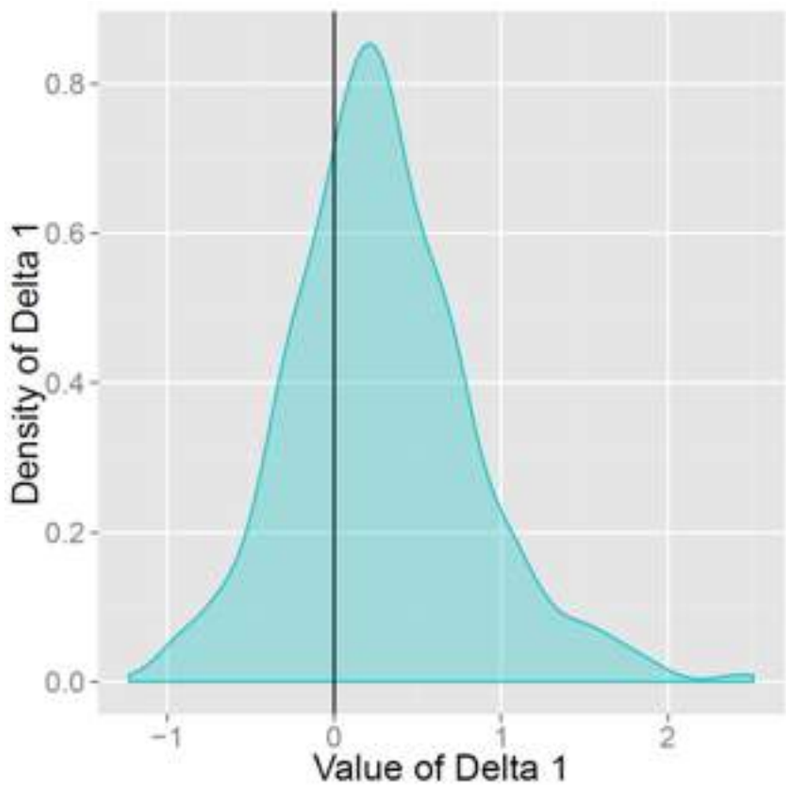
2. **Sensitive to therapy: Following a therapy (e.g., fluid bolus), does the score reflect decrease in severity?**

Task 2: Sensitive to changes in severity leading up to adverse events



$$\Delta_1 = \bar{s}_{0-6} - \bar{s}_{6-12}$$

Task 2: Sensitive to changes in severity leading up to adverse events

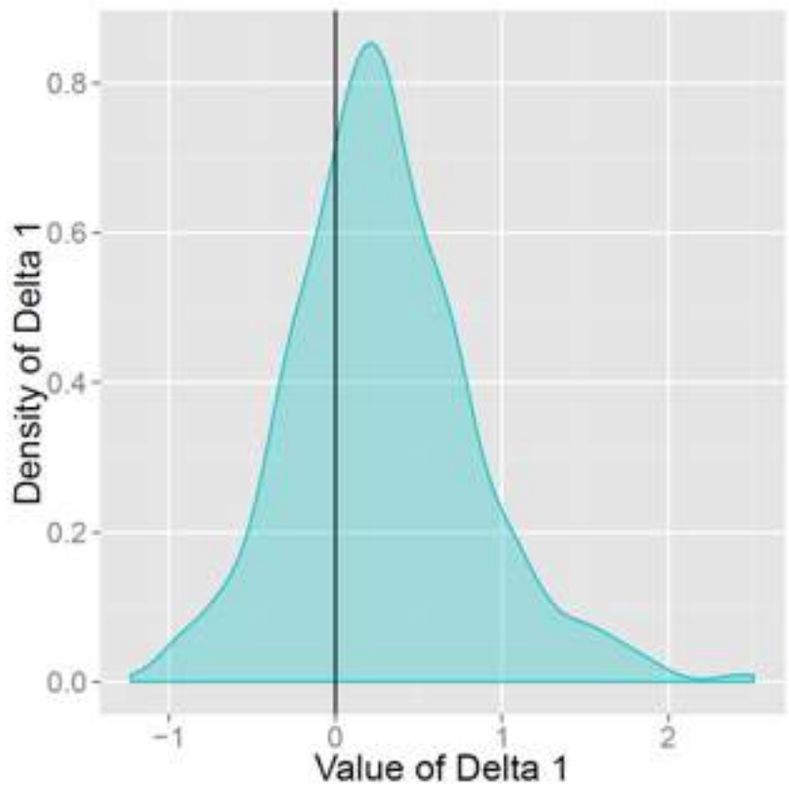


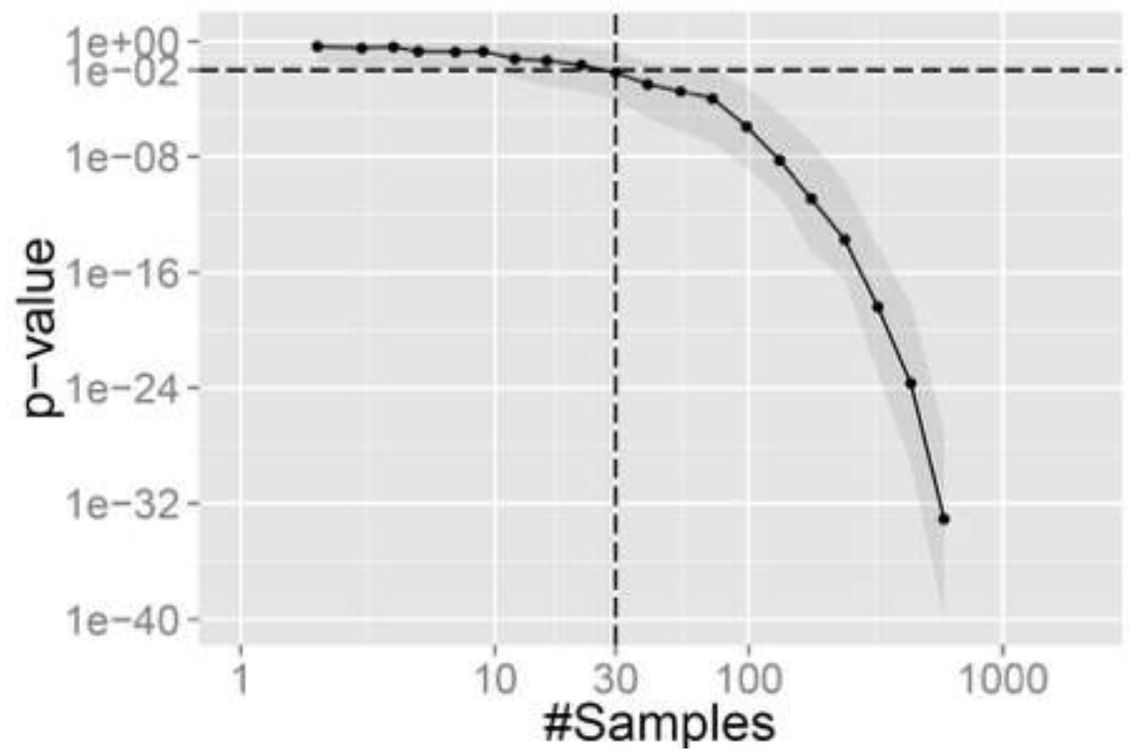
$$\Delta_1 = \bar{s}_{0-6} - \bar{s}_{6-12}$$

Paired-up

$\Delta_1 > 0$

Task 2: Sensitive to changes in severity leading up to adverse events





$$\Delta_1 = \bar{s}_{0-6} - \bar{s}_{6-12}$$

Type	Predictor	$p_{trend-up}$	# samples	Fraction of $\Delta_1 > 0$
Proposed Scores	*L-DSS* λ_s^1	<1e-32	587	0.71±0.037
	L-DSS λ_s^2	<1e-34	587	0.70±0.036

Numerical Analyses

1. **Ranking performance: does the proposed score order for severity more accurately than existing scores?**

2. **How sensitive is the score to more granular changes in severity beyond coarse grading for the stages of sepsis?**

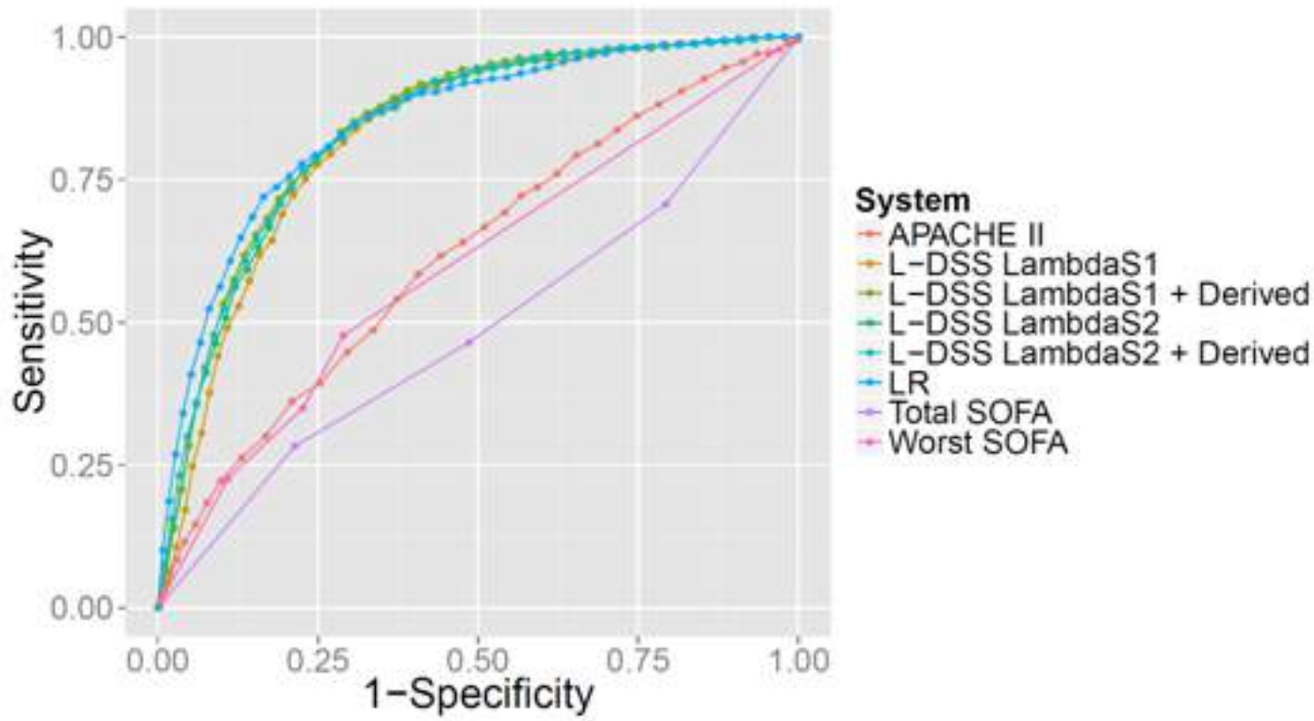
1. **Sensitive to impending adverse events: Leading unto an adverse event (e.g., septic shock), does the score reflect increase in severity?**

- **Predictive performance: Measure prediction performance of the resulting score.**

2. **Sensitive to therapy: Following a therapy (e.g., fluid bolus), does the score reflect decrease in severity?**

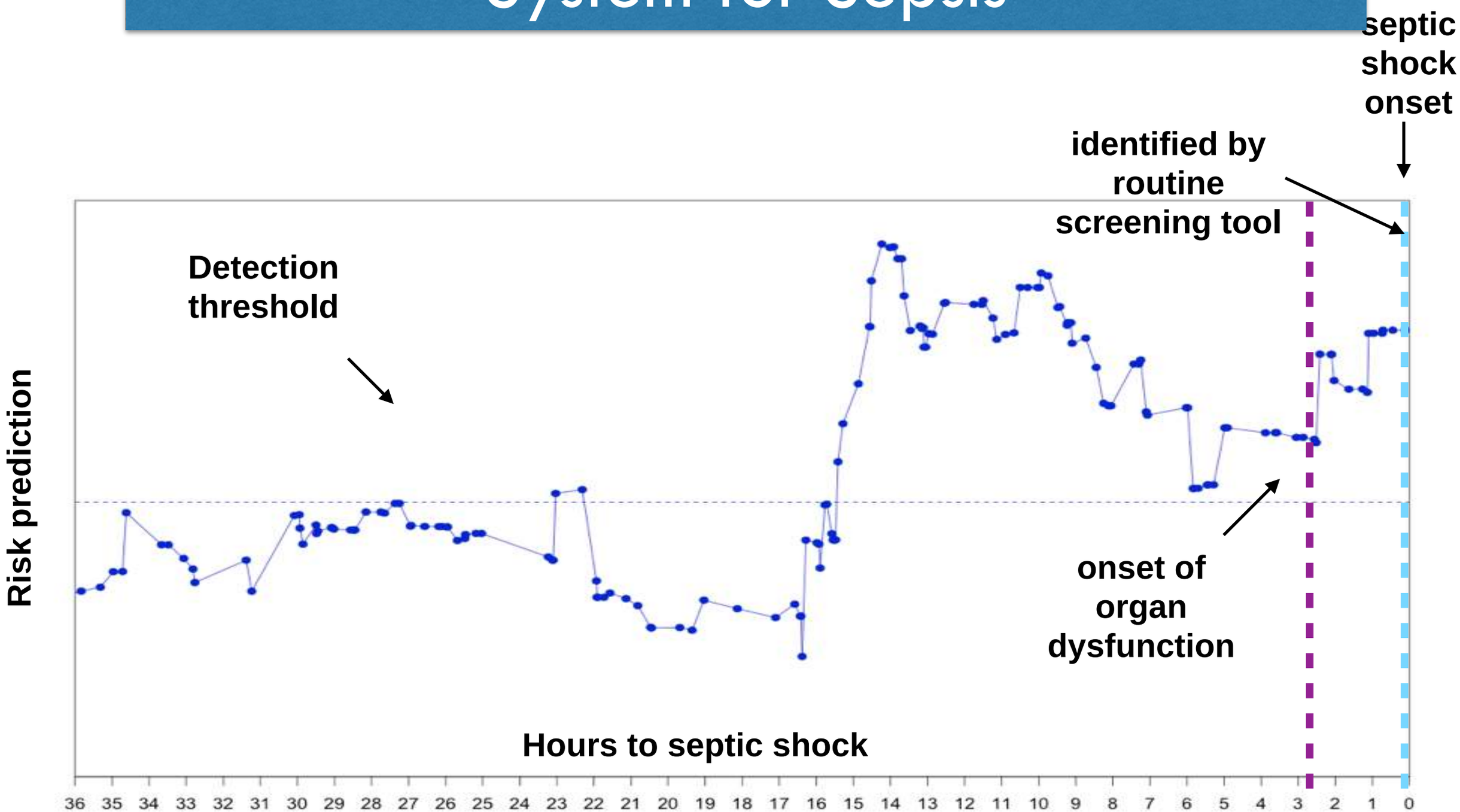
Early Detection of Septic Shock

- Develop an alert to identify those at risk for septic shock
- Raw DSS itself
- Raw DSS + 7 derived features — average and trend related features over a window



Type	Predictor	λ_{smooth}	Mean AUC for ED	95% CI
DSS-based predictor	*L-DSS*	λ_s^1	*0.836*	*0.824-0.849*
	L-DSS	λ_s^2	*0.853*	*0.841-0.865*
	L-DSS + Derived	λ_s^1	*0.856*	*0.844-0.868*
	L-DSS + Derived	λ_s^2	***0.857***	*0.844-0.868*
State-of-the-art clinical	*APACHE II*	-	*0.620*	*0.600-0.641*
	Worst SOFA	-	*0.601*	*0.581-0.621*
	Total SOFA	-	*0.477*	*0.453-0.500*
	Logistic Regression (LR)	-	***0.864***	*0.852-0.875*

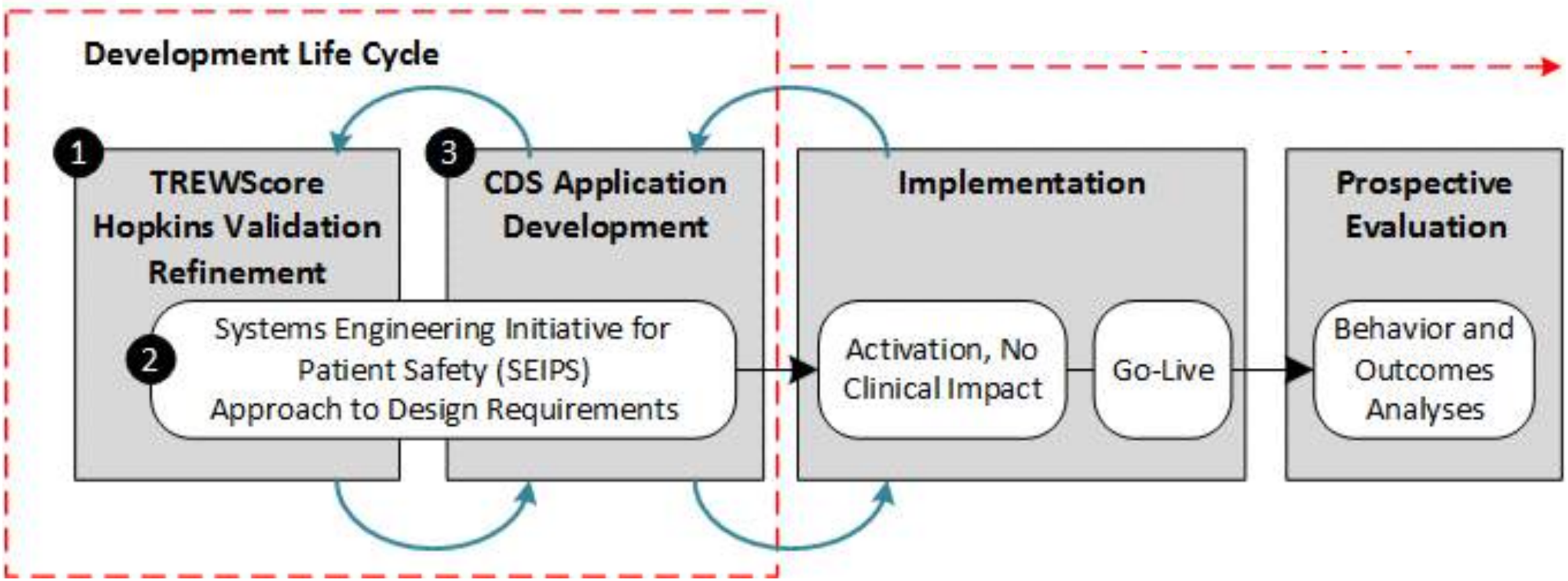
Targeted Real-time Early Warning System for Sepsis



- At 85% sensitivity, median of **25.2 hours** (IQR: 8.69-93.4) prior to shock onset
- **2/3rd** detected prior to any sepsis-related organ dysfunction
- Lower false alarm rate, earlier detection than routine screening protocol.

Pilot @ Hopkins integrated within EPIC

Use SEIPS to develop a deep understanding of workload and workflow —> Design Specification of the CDS



Numerical Analyses

1. **Ranking performance: does the proposed score order for severity more accurately than existing scores?**

2. **How sensitive is the score to more granular changes in severity beyond coarse grading for the stages of sepsis?**

1. **Sensitive to impending adverse events: Leading unto an adverse event (e.g., septic shock), does the score reflect increase in severity?**

   - **Predictive performance: Measure prediction performance of the resulting score.**

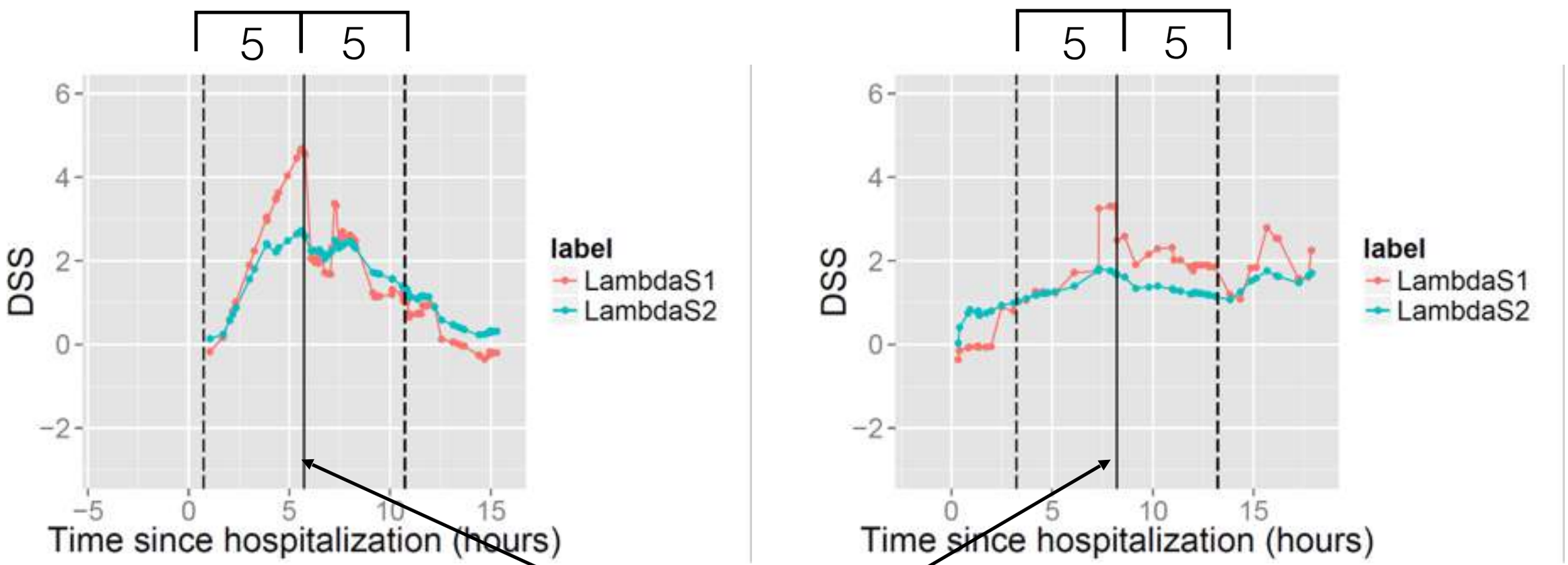
2. **Measure Responsiveness to therapy: Following a therapy (e.g., fluid bolus), does the score reflect decrease in severity?**

Responsive to therapy?

Is DSS responsive to therapy?

- construct validity: consistent with clinical expectations

Identified individuals in validation set who received bolus as treatment for sepsis (i.e. were septic, hypotensive (BP < 100), and received fluids of 30*body weight)



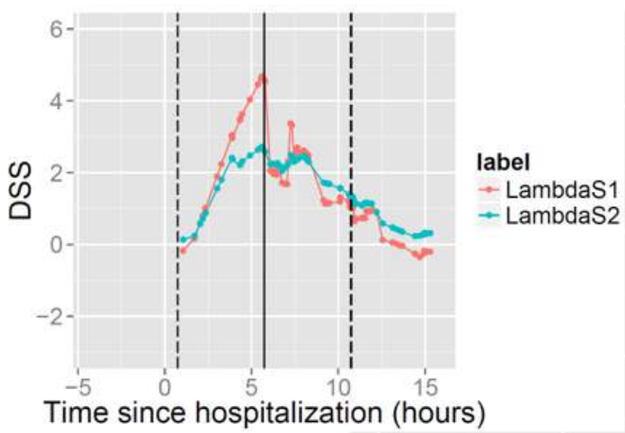
All solid lines mark interventions

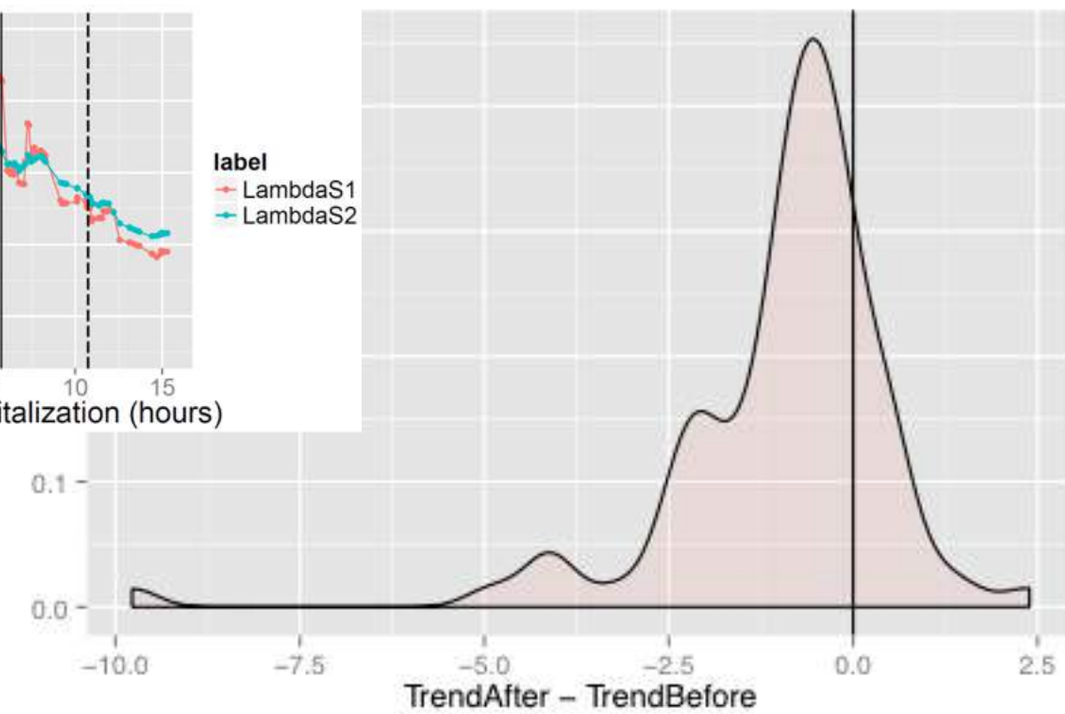
Changes in score post therapy

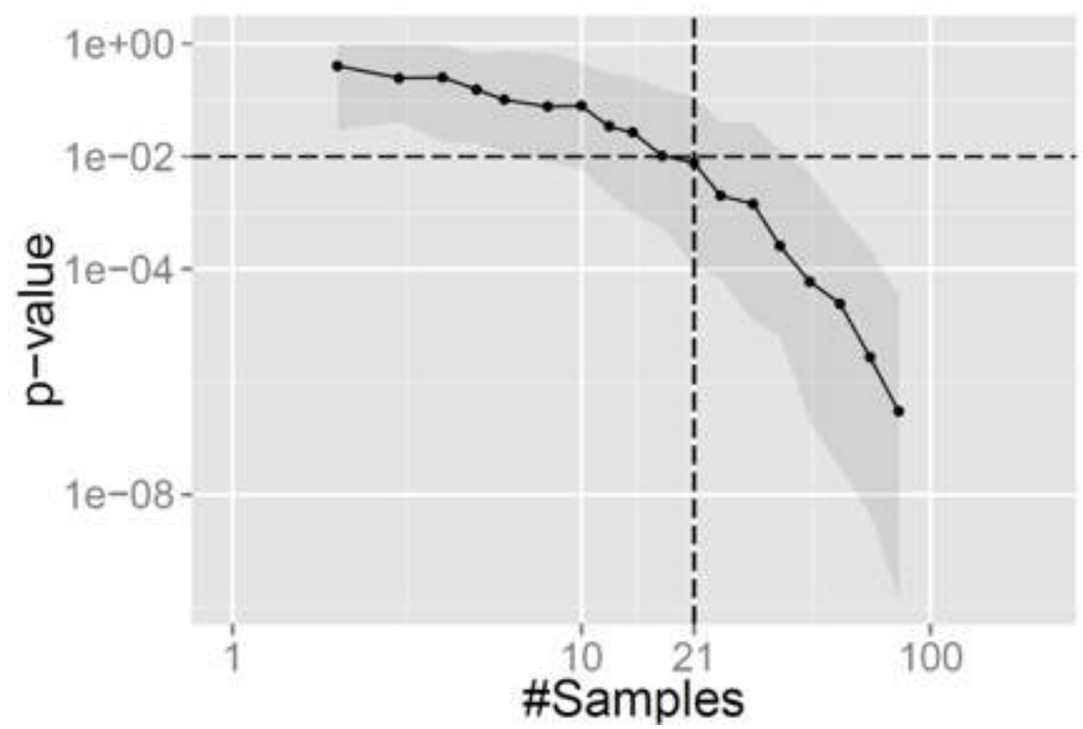
Pre/Post analysis:

- Comparing the trend of the DSS in the 5 hour period **post** treatment to the trend in the 5 hour period **pre** treatment.

- Expectation: good score should exhibit a negative change in the trend: upward trend is slowed down or even reversed.







Type	Predictor	λ_{smooth}	$p_{treatment}$	Number of examples	**Fraction of desirable (negative) trend changes**
Proposed Scores	*L-DSS*	λ_s^1	<2e-10	81	0.77±0.08
	L-DSS	λ_s^2	<5e-7	81	0.84±0.08

Learning Severity Scores and Risk Trajectories from Data

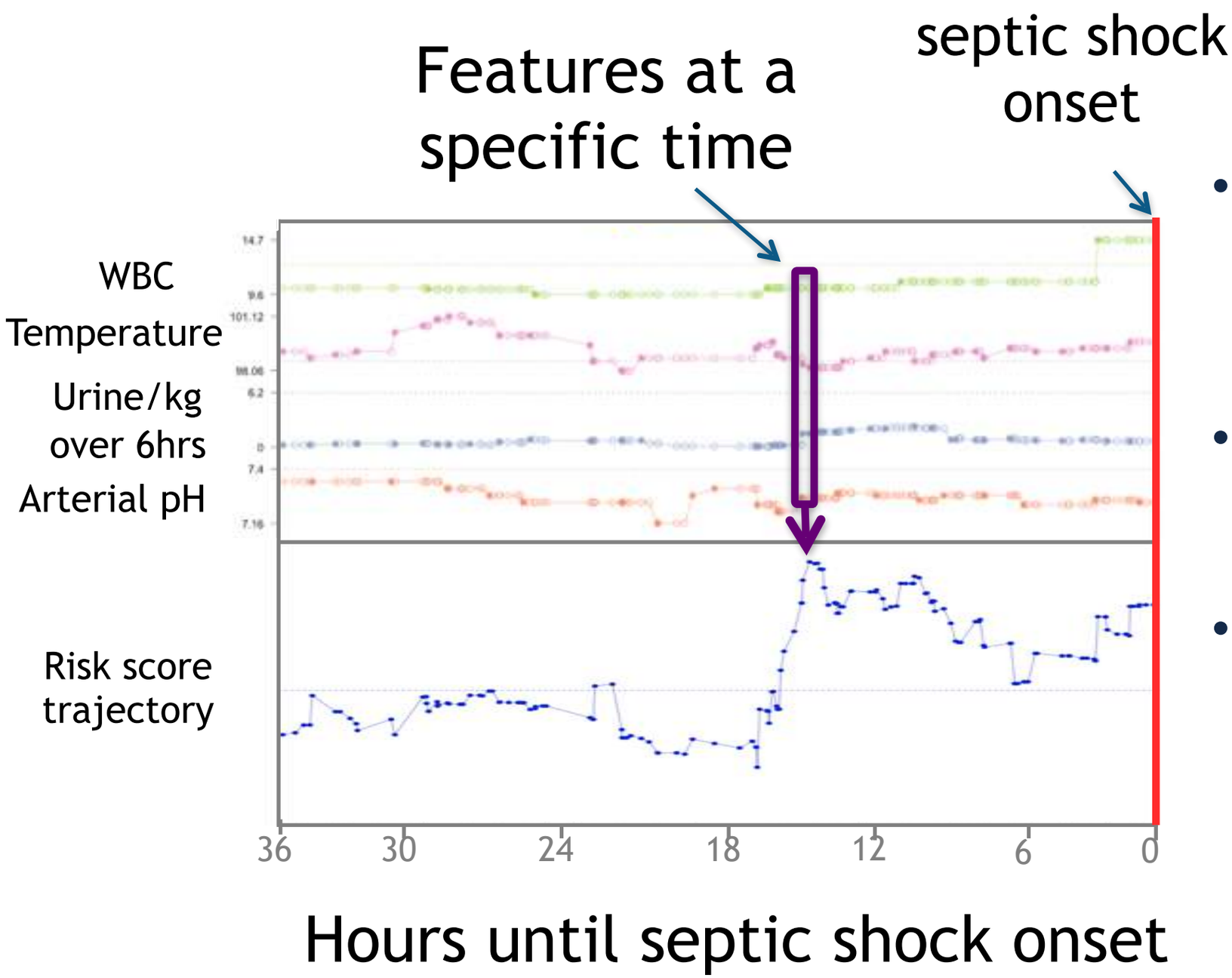
- **New framework for estimate disease severity**
  - Clinical comparisons: Ordering of disease severity states
  - Avoids bias due to interventional confounds (e.g. Paxton et al., 2013).

- Learned scores follow "construct validity"
  - Orders severity states with high accuracy
  - Sensitive to changes in severity leading up to adverse events
  - Sensitive to changes in severity post therapy

Clinical & Informatics
- High Early Detection performance
- Measure therapy responsiveness

Informatics & ML
- Explore other criteria for generating clinical comparisons
- Active learning with user in a loop
- Tailor to subpopulations and the individual

Risk-Monitoring



- **Vital signs:** heart rate, respiratory rate, temperature, Glasgow coma score…
- **Laboratory results:** BUN, creatinine, WBC, hematocrit…
- **Clinical history:** presence of chronic conditions, age…

Routinely collected measurements

Test	Measurement	Routinely Collected	Cost (USD)	Wait time (minutes)	Measurement time (minutes)
Vitals and Clinical History	Age, Weight	Y	$0	0	0
	Temperature	Y		0	0
	Heart rate	Y		0	0
	Heart Rhythm	Y		0	0
	Respiratory rate	Y		0	0
	Non-invasive blood pressure	Y		0	0
	Blood oxygen (SpO2)	Y		0	0
	FiO2	Y		0	0
	Riker Score	Y		0	0
	Glasgow Coma Scale	Y		0	0
	On dialysis	Y		0	0
	On pacemaker	Y		0	0
	Admission Diagnoses	Y		0	0

Free but are not accurate.

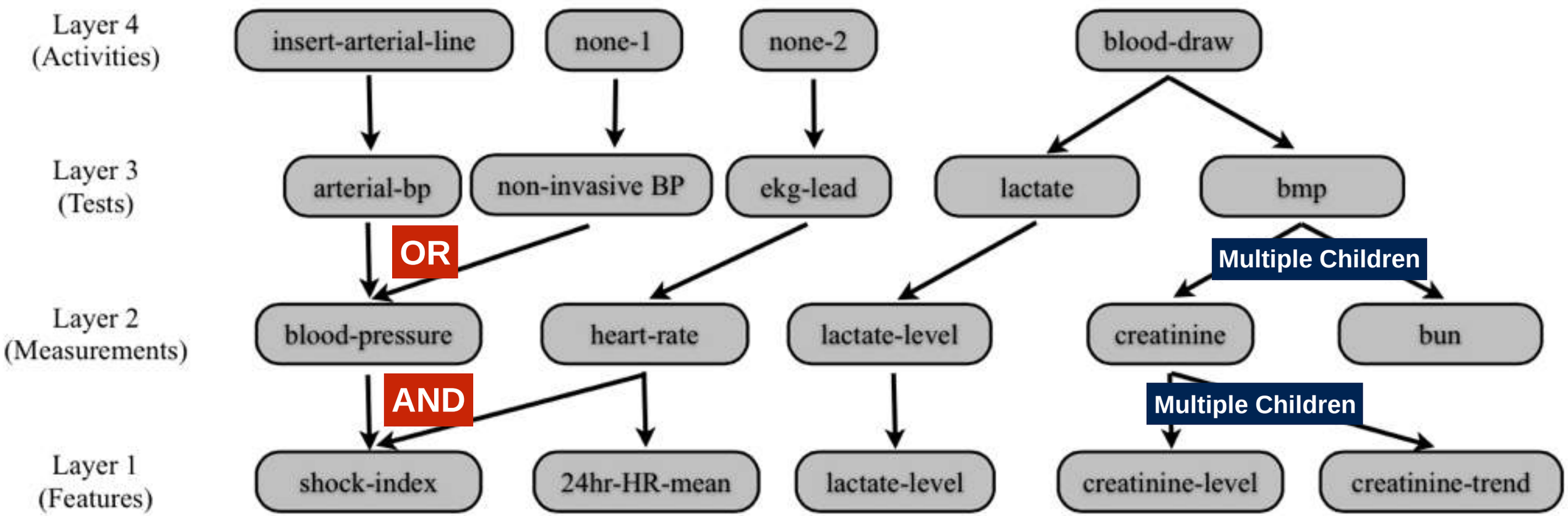
On-demand measurements

Test	Measurement	Routinely Collected	Cost (USD)			Wait time (minutes)	Measurement time (minutes)
Complete blood count (CBC)	White blood cell count	OD	CBC $24			50 mins	20 mins of nursing time for a blood draw
	Red blood cell count	OD					
	Platelets	OD					
	Hemoglobin	OD	$35				
	Hematocrit	OD	$9				
Metabolic Panel	Blood urea nitrogen (BUN)	OD	$14	BMP $24	CMP $30		
	Creatinine	OD	$14				
	Glucose	OD	$11				
	Sodium	OD					
	CO2	OD					
	Potassium	OD					
	Calcium	OD					
	Bilirubin	OD					
Blood Gases	PaCO2	OD	Blood Gas $72			50 mins	0
	PaO2	OD					
	pH	OD					

Others collected on demand with *varying types of costs*

Cost dependency Graph

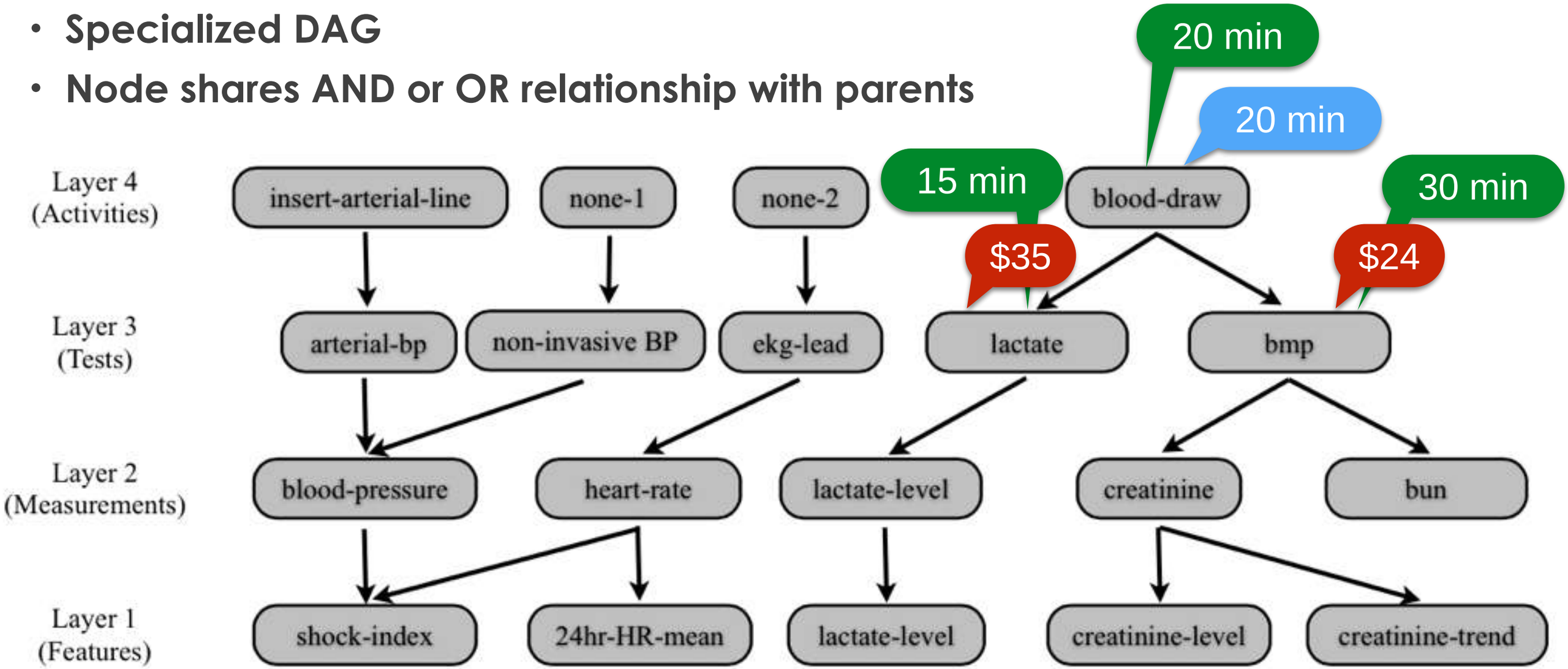
- **Specialized DAG**
- **Node shares AND or OR relationship with parents**



- **Tests generate multiple measurements**
  - **AND**: Measurements may require more than one test
  - **OR**: More than one test may generate a measurement
  - Similarly, features share AND or OR relationship with measurements

Cost dependency Graph

- **Specialized DAG**
- **Node shares AND or OR relationship with parents**



Staff-time — **Costs at the activity level**

Financial Cost — **Costs at the test level**

Wait time — **Costs at the activity and test level**

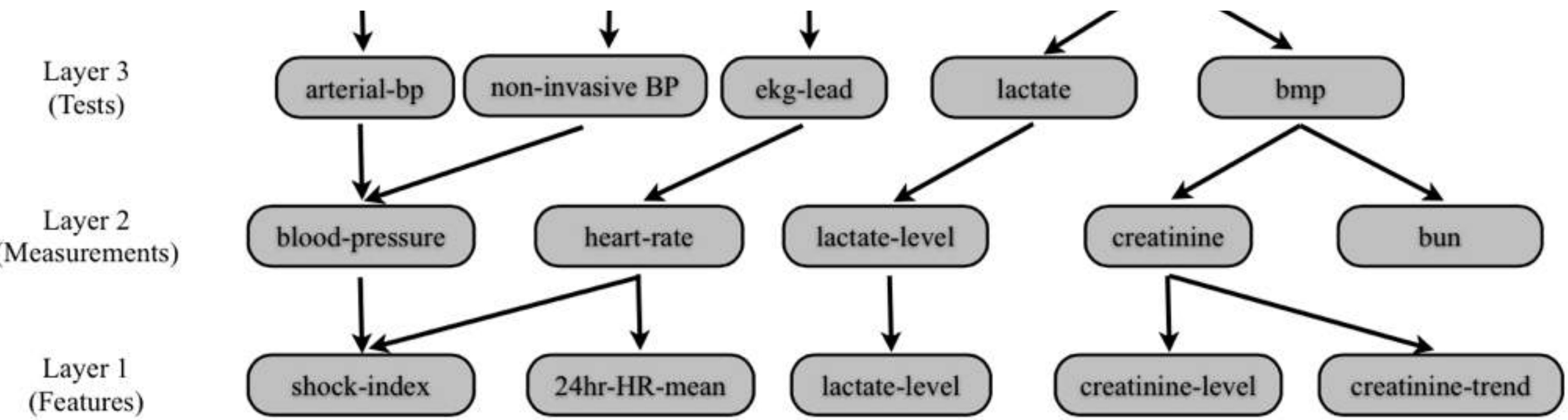
Cost-Sensitive Learning

- **Empirical Risk Minimization**

$$\underset{\beta \in \mathbb{R}^n}{\text{minimize}} \quad f(\beta) := \frac{1}{N} \sum_{i=1}^{N} L\big(\beta; (x_i, y_i)\big) + R(\beta)$$

Loss **Regularizer/ Penalty**

- **Goal: Construct Regularizer given the Cost-Dependency Graph**



Cost-Sensitive Learning

- Classifier Cascades:
  - Viola and Jones, IJCV 2004; Raykar et al., 2010

- Incremental Cost-Sensitive Classification:
  - MDPs - Ji and Carin, 2007
  - Weiss, Sapp, Taskar 2013

- Test-time Budgets:
  - Xu, Kusner, Chen and Weinberger 2013
  - Xu, Weinberger and Chapelle, 2012

- Structured Sparsity:
  - Group Regularizers (e.g., Bach et al.)
  - Tree and Forest Regularizers:
  - OSCAR (Bundel et al.,)

Majority of works on cost-sensitive classification assume costs given at the feature level

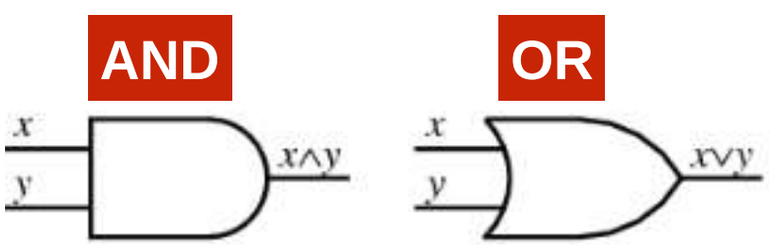
Structured sparsity work do not show how to obtain the form of the structured regularizer.

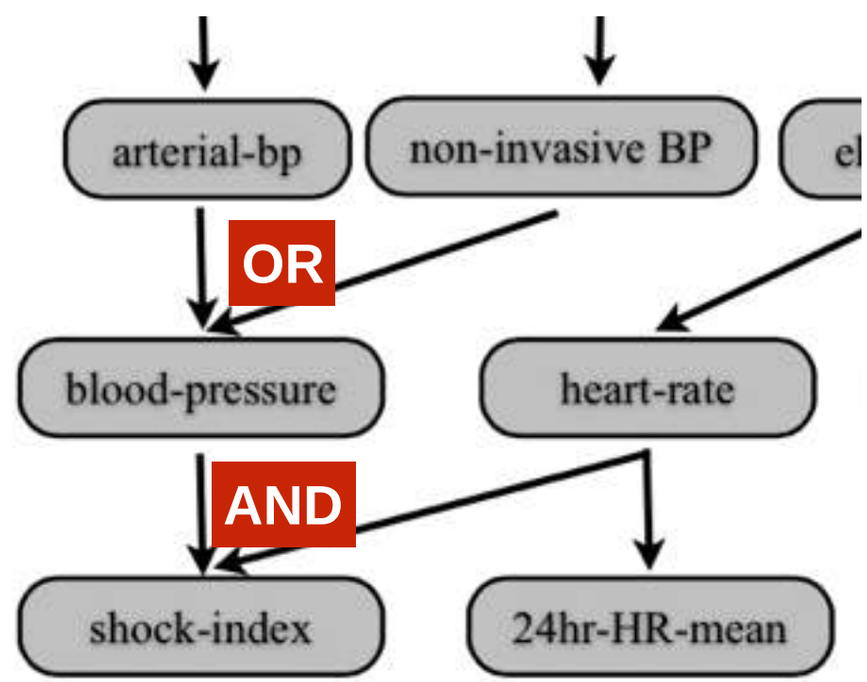
Cost dependency Graph

REMARK 1: Our graph is a multi-layer Boolean Circuit.

Reminder: Boolean circuits

- Circuit defined as a finite DAG.
- Each vertex corresponds to either a basis function or one of the inputs
- Example basis functions include AND, OR, NOT functions



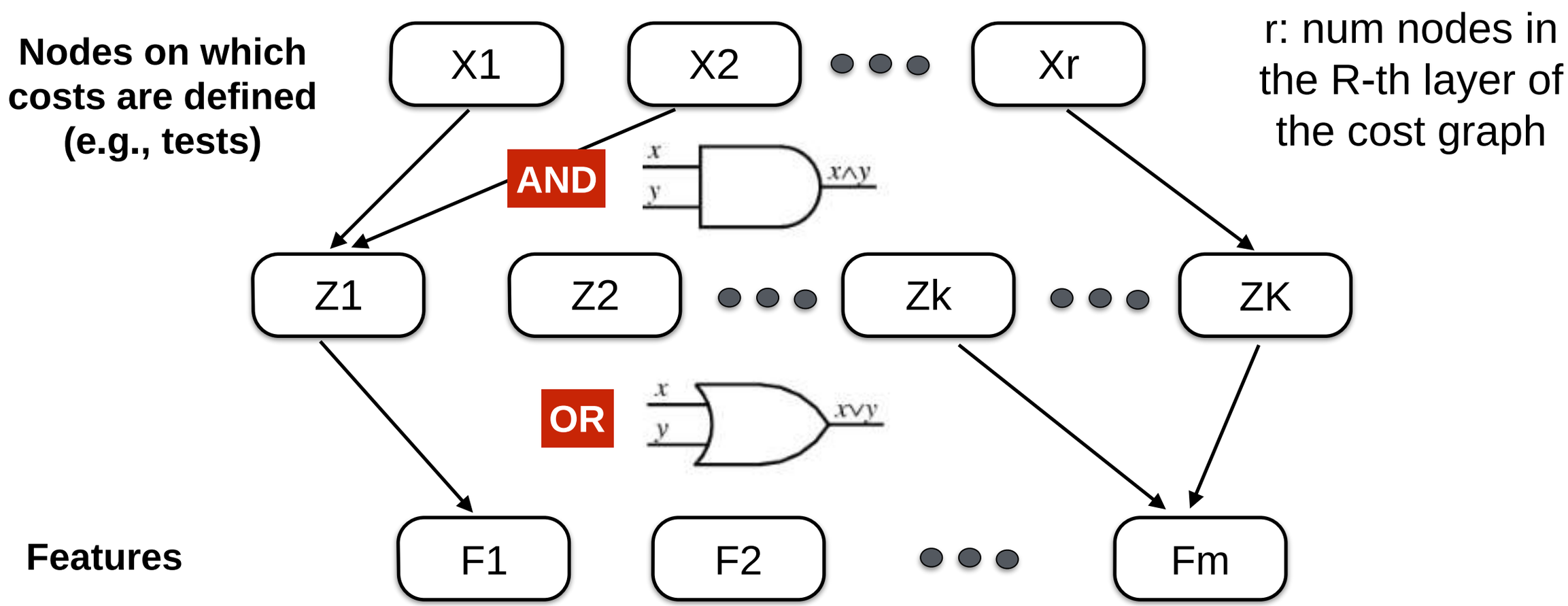


Cost dependency Graph

REMARK 1: Our graph is a multi-layer Boolean Circuit.

REMARK 2: By removing double negations, applying the laws of distribution, and using De Morgan's laws, the r-layer circuit may be reduced to a 3-layer boolean circuit in disjunctive normal form

(e.g., B. Pfahringer. Conjunctive normal form. In Encyclopedia of Machine Learning, Springer, 2010)

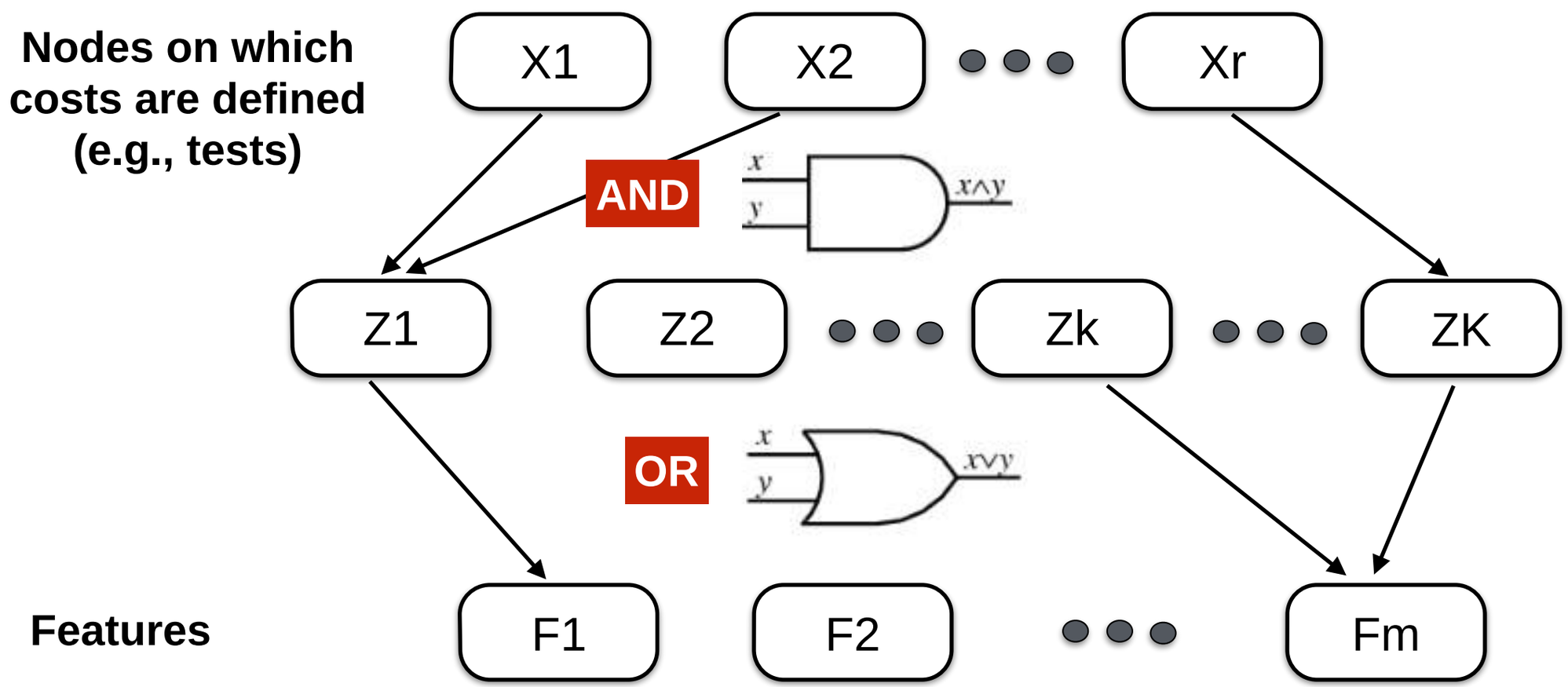


Cost dependency Graph

$$\mathcal{R}_{relax}(\vec{\beta}) = \lambda \sum_{k=1}^{r} \mathcal{C}_k || \vee_{j \in \mathcal{S}_k} \beta_j ||_\infty$$

$\mathcal{C}_k$ Cost associated with the kth input (e.g., test)

$\mathcal{S}_k := \{j : g_{z_j}(.)$ depends on the logical values of Xr $\}$



Sepsis: Our Target Condition

High mortality and morbidity

Septicemia is the 11th leading cause of death in the US

- Over 750,000 cases of severe sepsis or septic shock annually and rising
- Mortality of septic shock estimated between 30-60%
- Patients with sepsis have increased hospital stays and long-term morbidities

High financial cost

- $15.4 billion dollars spent nationally

Mortality and length of stay decreased with timely treatment [Kumar et al. 2011]

- For every hour that antibiotic treatments were delayed, risk of mortality went up by 7.6%

P. R. Dellinger et al. Critical Care Medicine, vol. 41, no. 2, pp. 580–637, 2013
D. C. Angus et al., Critical Care Medicine, vol. 29, no. 7, 2001.
G. Kumar et al., Chest, vol. 140, no. 5, pp. 1223–1231, 2011.
HCUP Facts and Figures: Statistics on Hospital-based Care in the United States, 2009.

Dataset and Population

7 years worth of data from 4 ICUs in an academic medical center

MIMIC II Clinical Database

- **Electronic health records (EHR) collected from intensive care units as the Beth Israel Deaconess Medical Center from 2001 to 2008**

Study population

- **16,232 adult patients included (age ≥ 15 years)**
- **60% Training set**
- **40% Test set**

A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic sig- nals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13).
R. Dellinger *et al.* "Surviving sepsis campaign: International guidelines for management of severe sepsis and septic shock," *Critical Care Medicine*, vol. 41, no. 2, pp. 580–637, 2013.

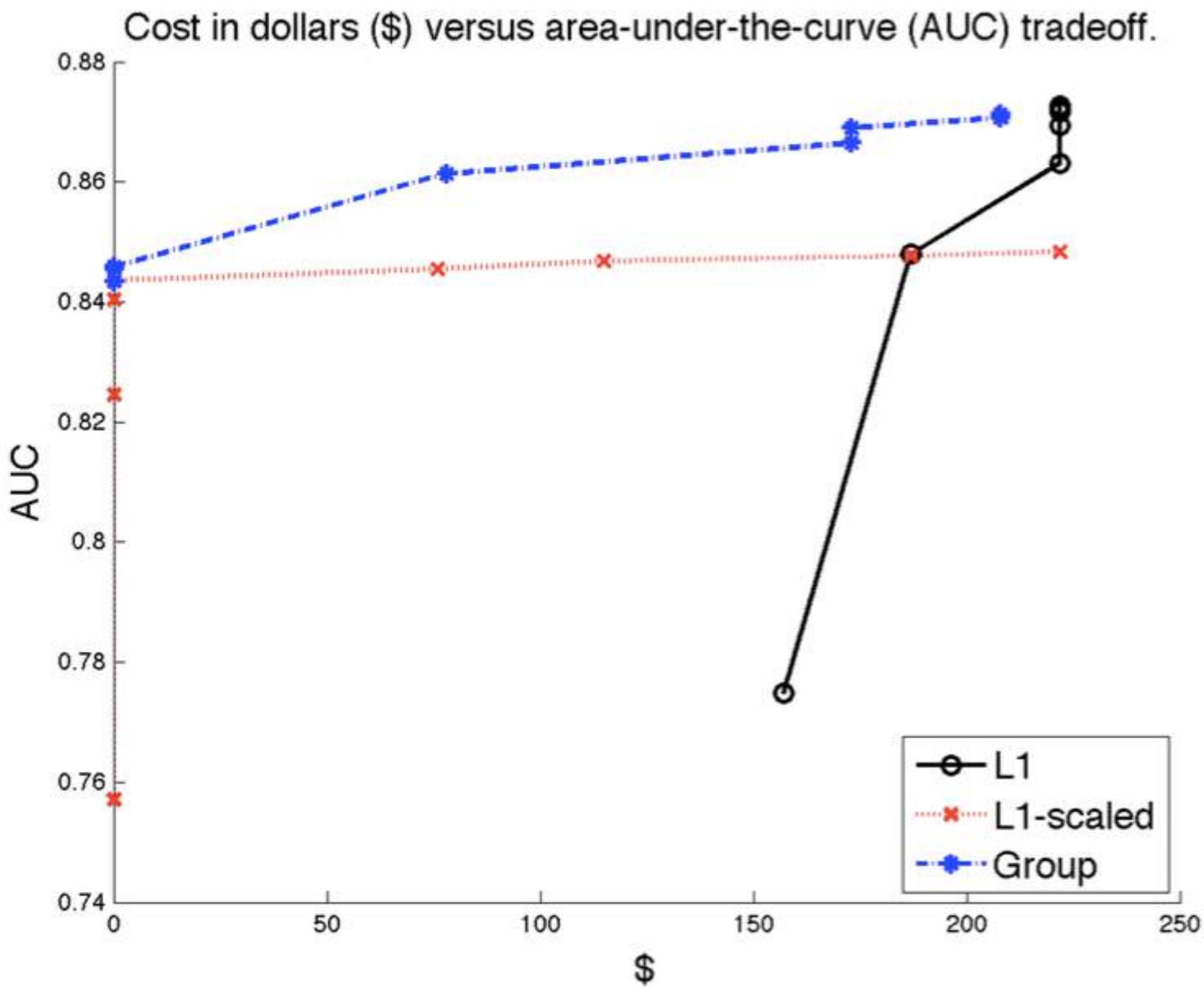
Range of Models Recovered

Models	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
Sensitivity at 0.85 specificity	0.61	0.66	0.65	0.72
AUC	82.79 ± 0.55	84.45 ± 0.64	84.75 ± 0.55	87.21 ± 0.46
Financial Cost	\$0	\$0	\$72	\$168
Caregiver Time	0 minutes	10 minutes	0 minutes	30 minutes
Result Time	0 minutes	10 minutes	50 minutes	50 minutes
Tests Required	routine	routine, urine	abg, routine	abg, cbc, cmp, hct, hemoglobin, routine, urine
Activities Required	none	urine	arterial stick	arterial stick, blood draw, urine

- Cheapest model:
  - **Sensitivity of 0.61**
  - \$0 / no additional measurements / no wait time

- More expensive model
  - **Sensitivity of 0.72**
  - \$168 / 30 mins of nursing time / 50 mins wait time

Naive accounting of cost



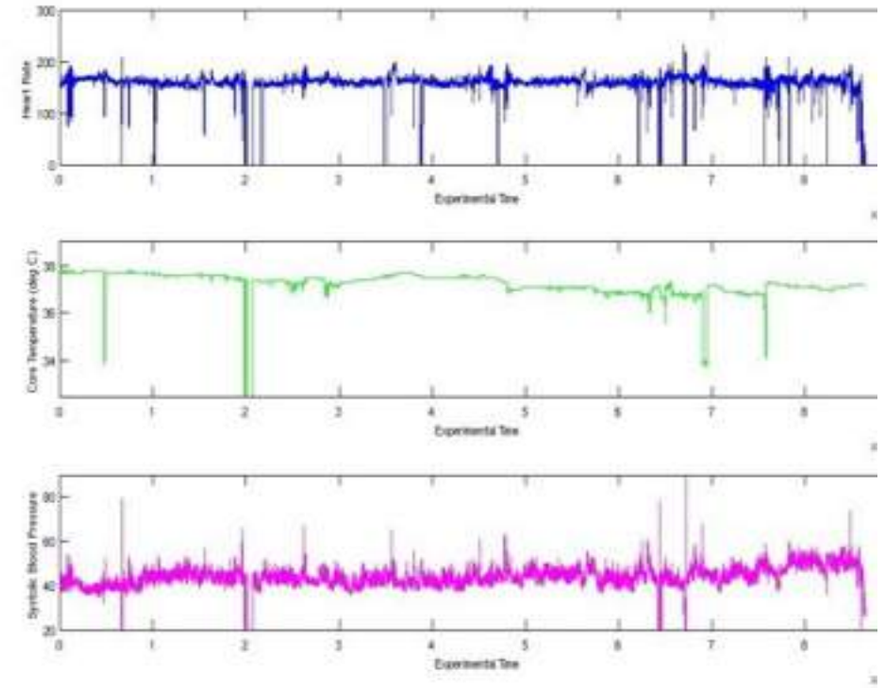
Electronic Health Data

Genomic data

Sensors & Devices

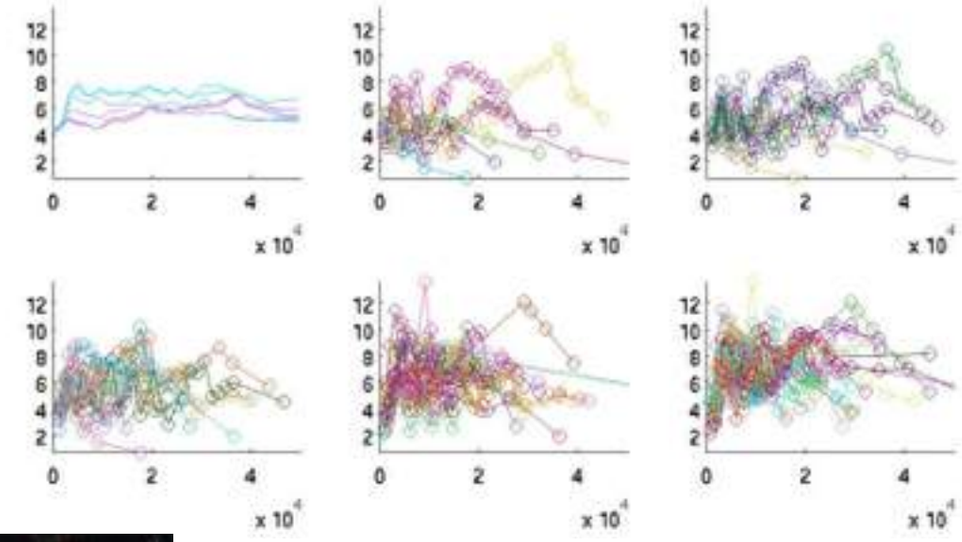
Administrative Claims

Progress notes



Continuous physiologic measurements

Imaging Data

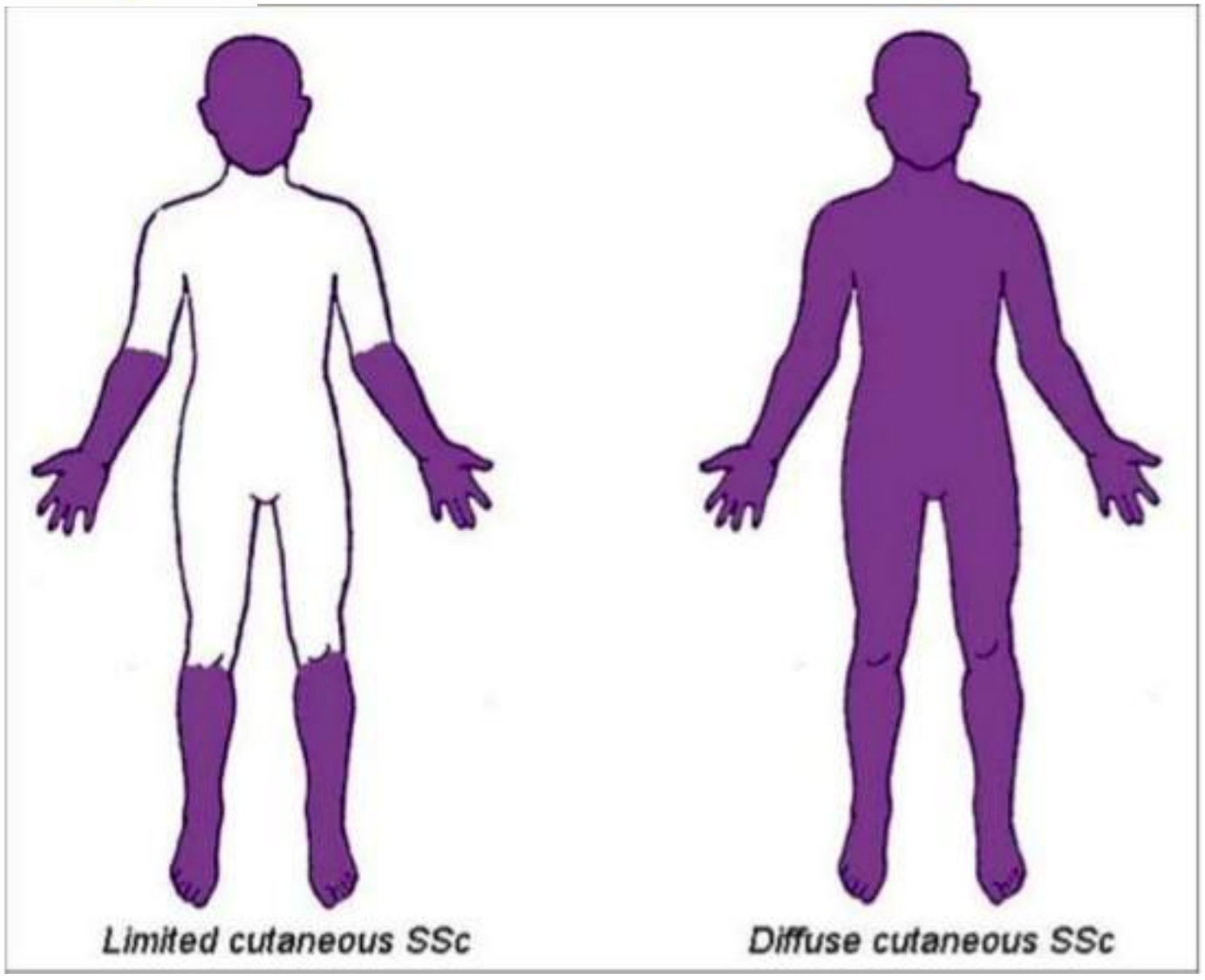


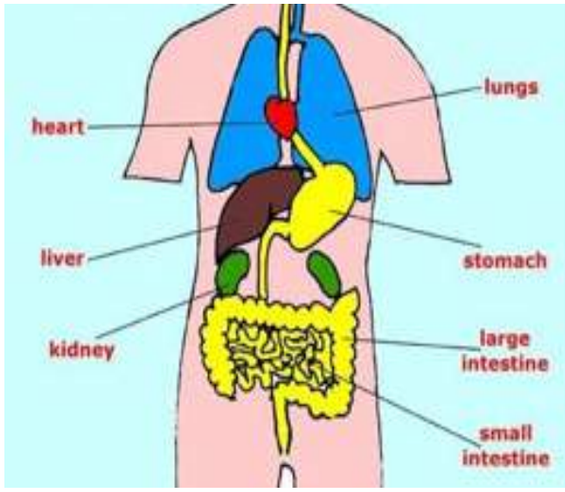
Discrete Events: Laboratory

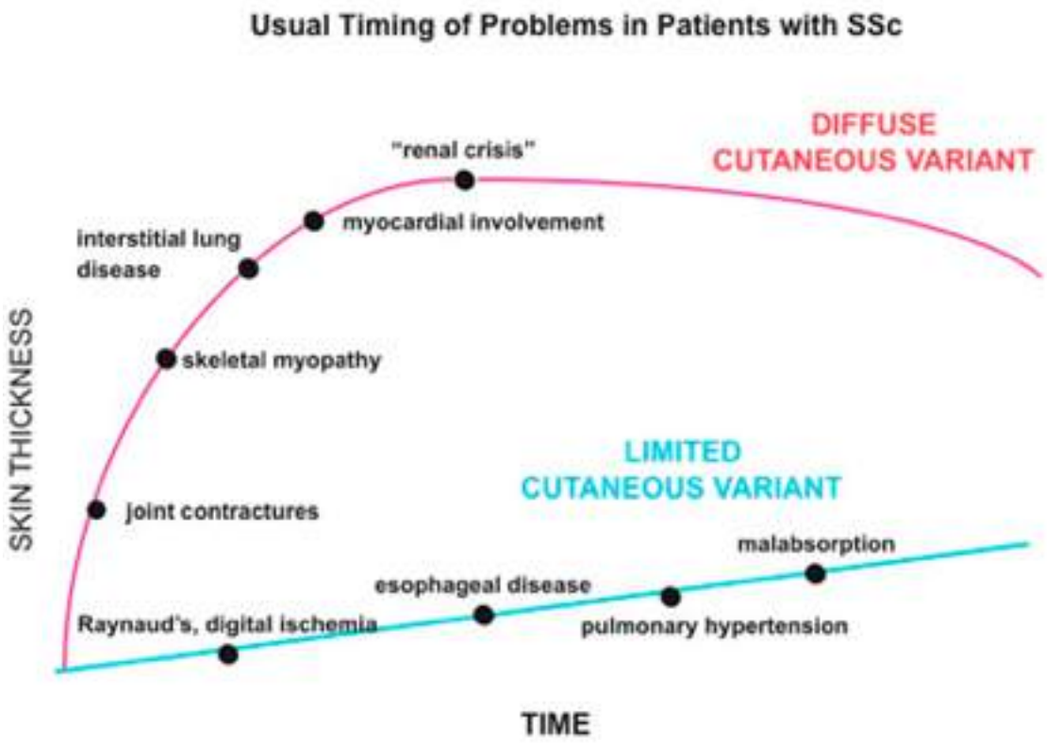
Interventions: Medicines, Procedures

Modeling variability across individuals in complex, chronic diseases: Autoimmune diseases

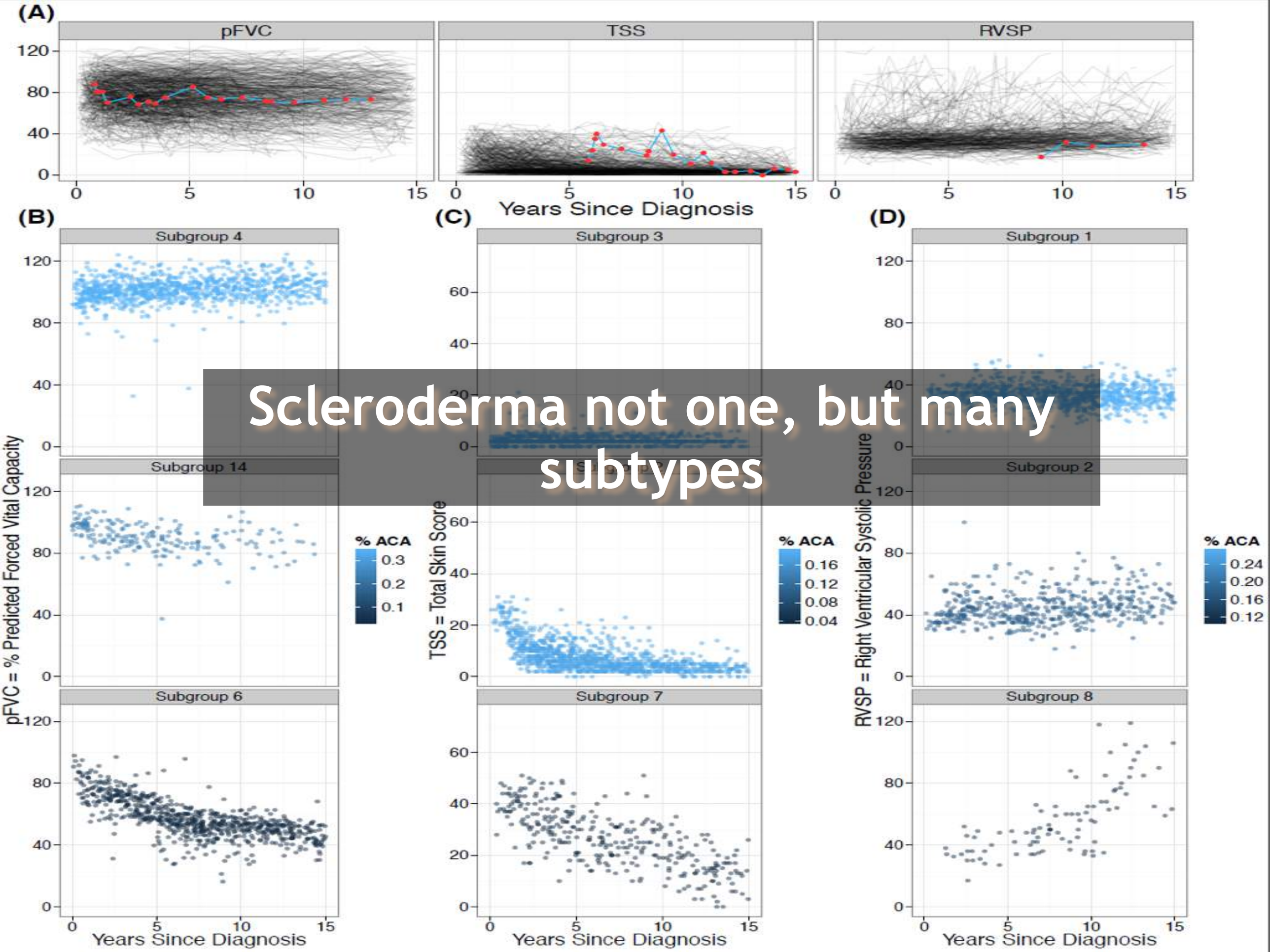
Smart and Connected Health







http://www.hopkinsarthritis.org/wp-content/uploads/2011/04/image-11.jpg
http://www.slideshare.net/maushard/skin-manifestations-of-scleroderma-by-dr-lorinda-chung-md

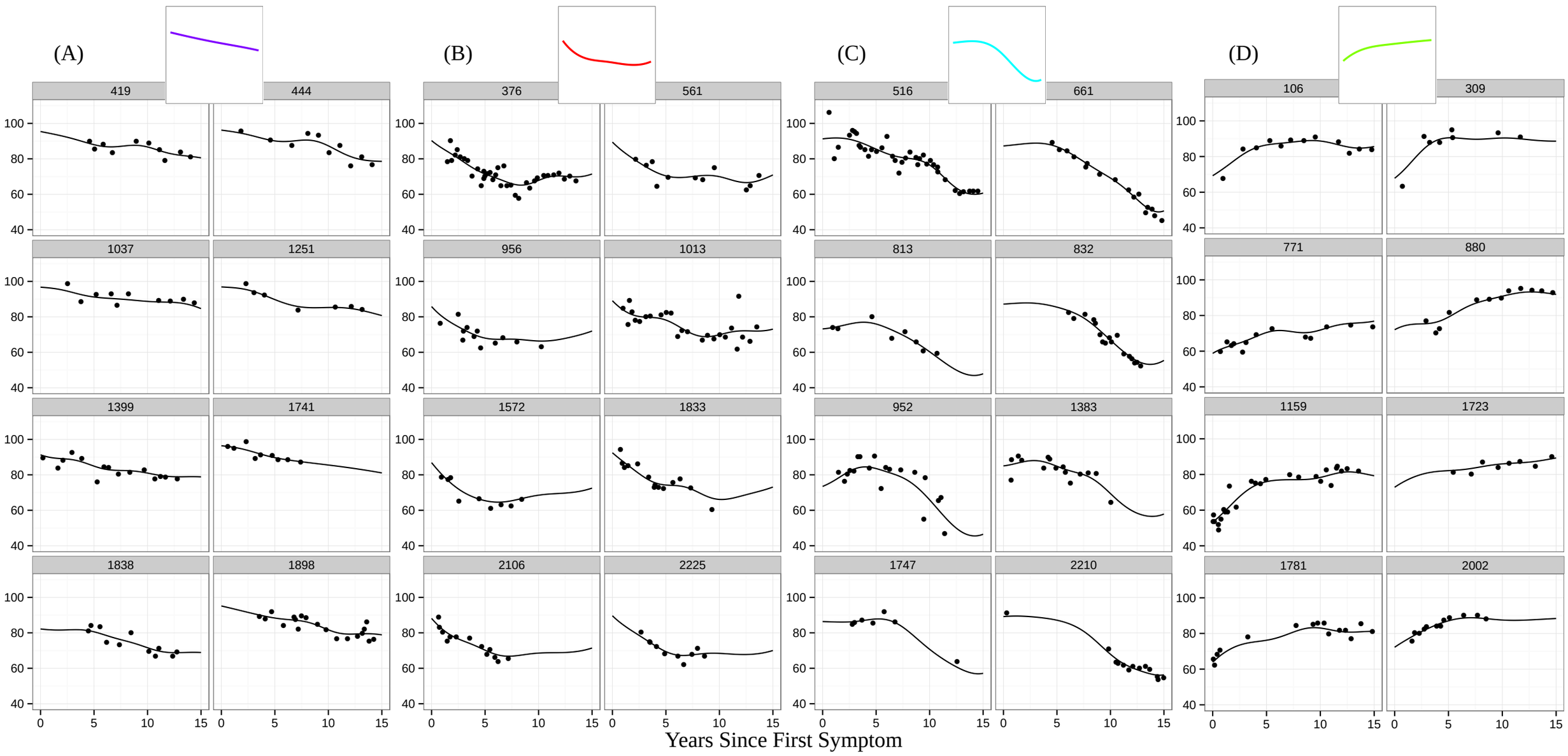


Scleroderma not one, but many subtypes

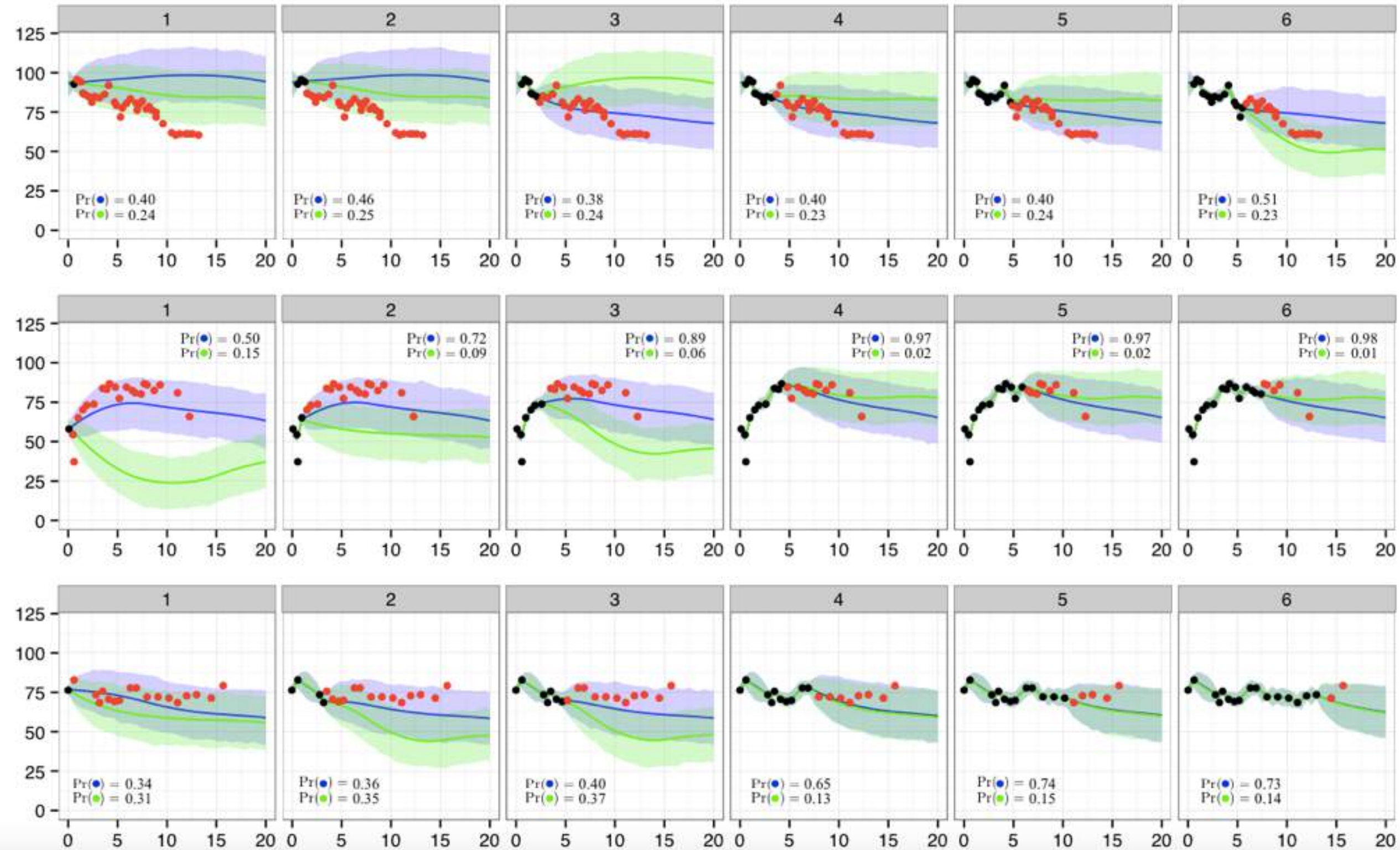
Subtyping from Irregularly, Sampled EHD

- *Subtypes* — groups of individuals with similar disease progression patterns.

Automated Subtype Discovery from data: Schulam, Wigley, Saria. AAAI 2015

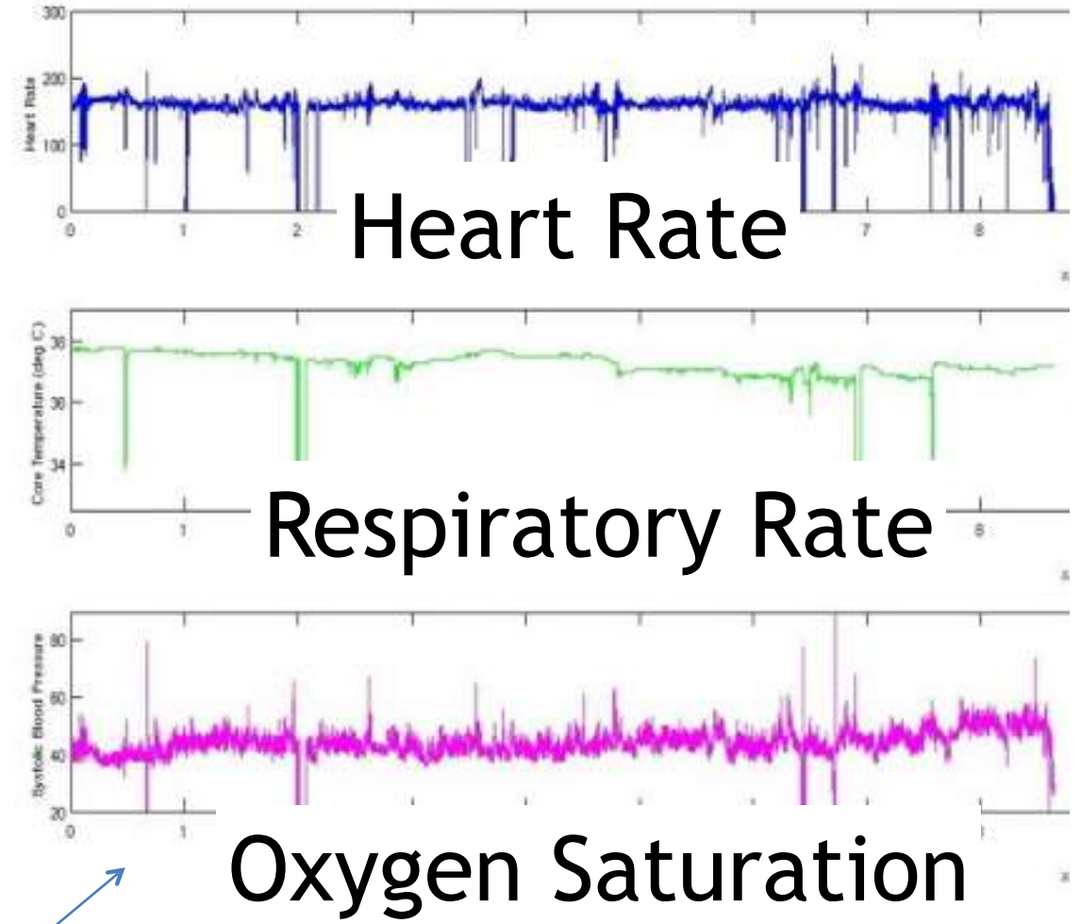


Prediction Models for Chronic, Multiphenotyic Diseases (e.g., predicting lung function in sclerdoderma)



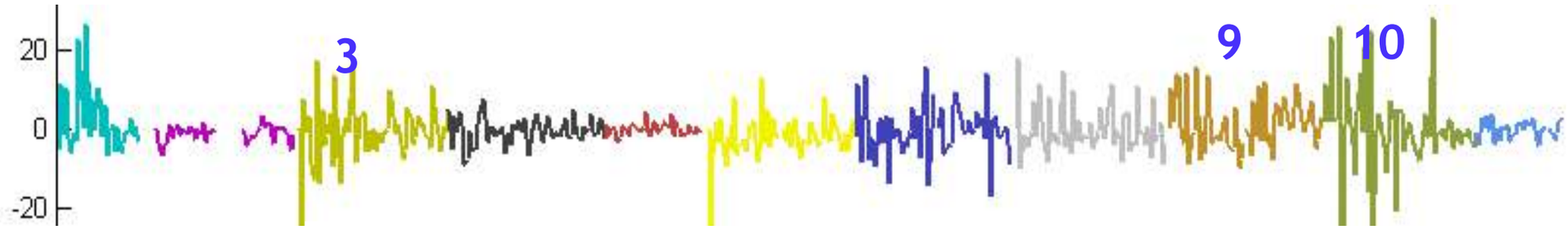
For ***each*** **individual, predict their disease course.** **(under review)**

Physiologic Signals: Collected but Dropped after 48 hours

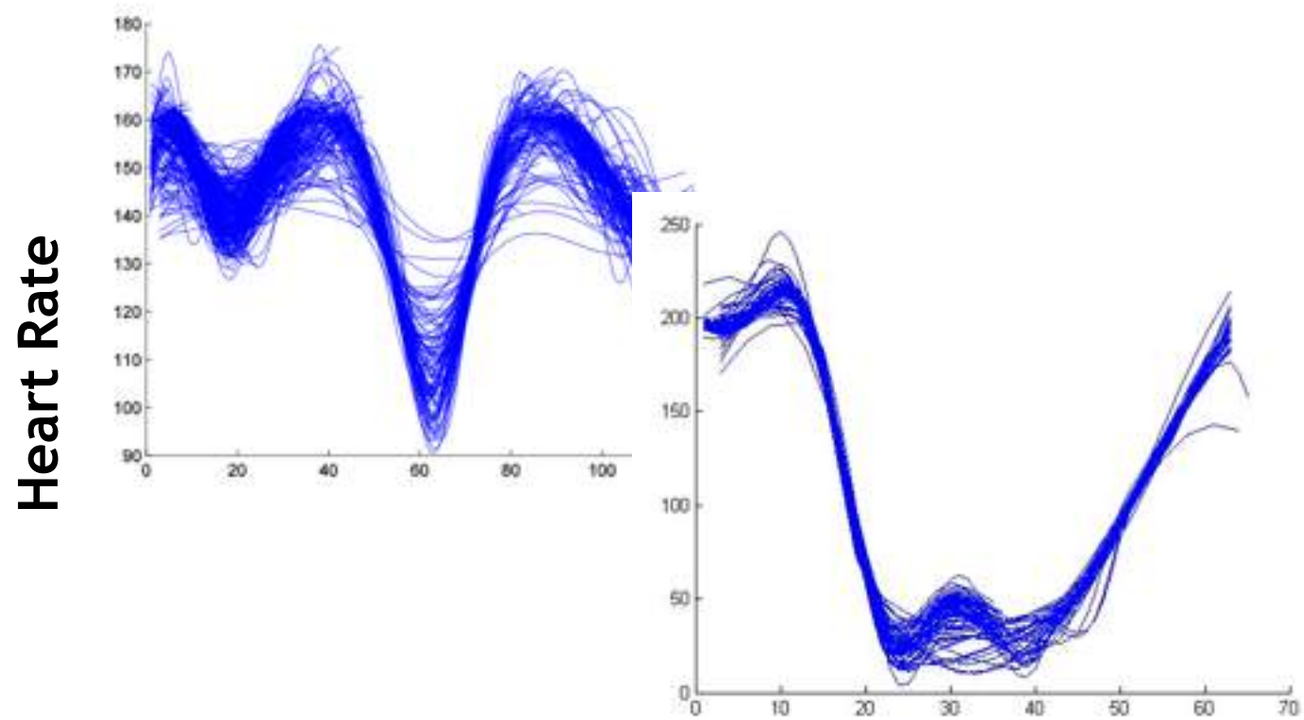


Computational Markers

Example signatures learned with TSTM are shown below:



$$x_t = a^i x_{t-1} + \varepsilon$$
$$\varepsilon \sim N(0, \sigma^i)$$



S. Saria, A. Duchi, D. Koller. **Learning Deformable Motifs in Continuous Time Series data**. International Joint Conference on Artificial Intelligence (IJCAI), 2011.

S. Saria, A. Rajani, J. Gould, D. Koller, A. Penn. **Integration of Early Physiological Responses Predicts Later Illness Severity in Preterm Infants**. Science *Translational Medicine*, September 2010. Vol. 2, Issue 48.

S. Saria, D. Koller, A. Penn. **Learning individual and population level traits from clinical temporal data.** In *Proc. Neural Information Processing Systems (NIPS), Predictive Models in Personalized Medicine workshop*. 2010.

Science AAAS

AAAS.ORG | FEEDBACK | HELP | LIBRARIANS

Daily News | Enter Search Term

ALERTS | ACCESS R

NEWS | SCIENCE JOURNALS | CAREERS | BLOGS & COMMUNITIES | MULTIMEDIA | COLLECTIONS

News Home | ScienceNOW | ScienceInsider | Premium Content from Science | About Science News

Home > News > ScienceNOW > September 2010 > New Model Predicts Complications in Preemies

Science Video Portal

Science NOW — UP TO THE MINUTE NEWS FROM SCIENCE

Science LIVE
Every Thursday at 3:00 p.m. EST
Upcoming:
New Treasures in the Genome
13 September

New Model Predicts Complications in Preemies

by Cassandra Willyard on 8 September 2010, 2:10 PM | Permanent Link | 0 Comments

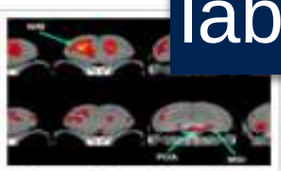
Email | Print | 0 | More | PREVIOUS ARTICLE | NEXT ARTICLE

Fully automated: uses routinely collected data available from bedside monitors. Less training required to use.

Non-invasive: Only relies on physiology. No need for additional lab tests.

RECENT ARTIC

Studies of Moto
Liver Transpl
Researchers 20
Award



ScienceShot: Bird Brains Show Their Fear of Faces

ScienceShot: The Bluest Fruit

however, the Apgar score is right only about 70%
models that factor in white blood cell count and
"they require a lot of invasive testing," says Anna
(LPCH) in Palo Alto, California.

The researchers, including Penn and co-senior a
set out to develop a more accurate yet noninvas
newborns. The researchers selected 138 infants
weighed less than 2 kg. The team classified the
developed. Babies in the high-risk group died or
and lung and heart problems. Babies in the low-
respiratory distress.

Next, the researchers examined physiological d

dicine

Events

Magazine

Newsletter

nication

determining premature infants' risk of illness

BY DIANNE KLEIN

Stanford University researchers have developed a revolutionary, non-invasive way of quickly predicting the future health of premature infants, an innovation that could better target specialized medical intervention and reduce health-care costs.

"What the PhysiScore does is open a new frontier," said Anna Penn, MD, PhD, an assistant professor of pediatrics at the School of Medicine and a neonatologist at Lucile Packard Children's Hospital. "The national push

Lucile Packard Children's Hospital

ed

eye or by conventional monitoring

Science *Translational Medicine*, 2010

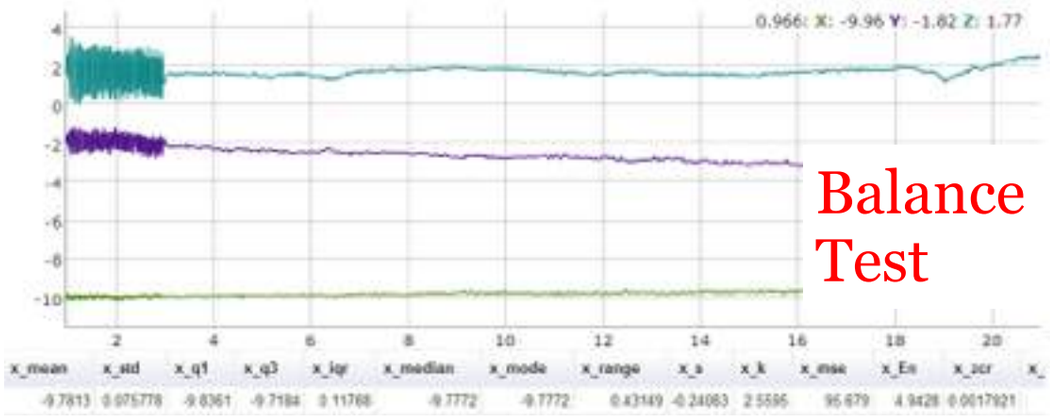
HopkinsPD: Automated Home-based Illness Severity monitoring

Largest database to date: 50K+ hours of data, 500+ individuals, 20+ countries

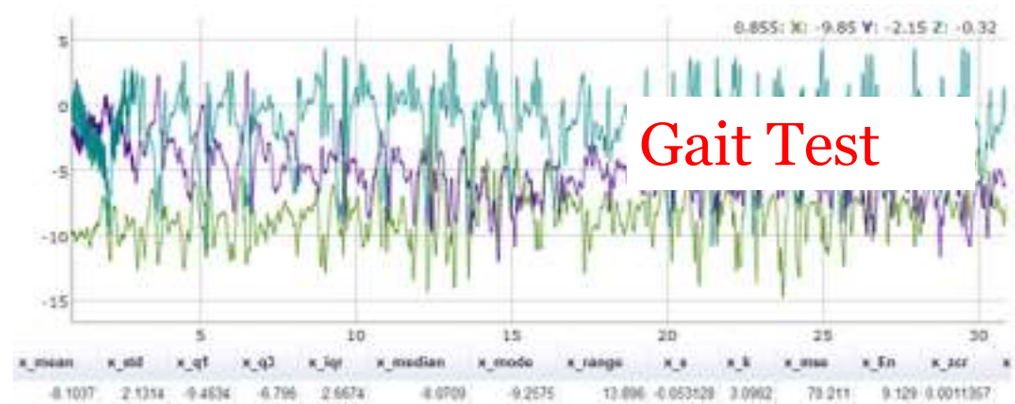
Voice Test

length	amp_mean	amp_std	amp_p1_1	amp_p1_0	amp_p2_2	amp_p2_1	amp_p2_0	amp_dfa	pitch
11.5	0.26566	0.043084	0.0093675	0.092364	0.001386	-0.041914	0.5502	1.4881	142.09

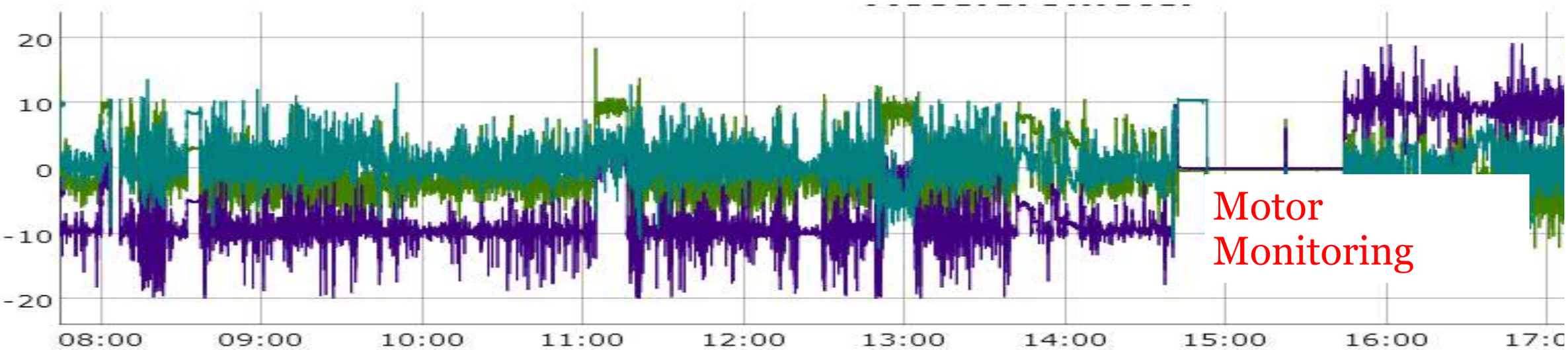
Location Monitoring



Balance Test



Gait Test



Motor Monitoring

Summary

Novel learning algorithms

- Current methods are highly susceptible to provider practice patterns.
- Interventional Confounds (ICs)
- New ranking based learning algorithm that avoids bias due to ICs.
- Application in the context of adverse event detection.

Extension to Practice-Cost Sensitive Learning

- Incorporate cost-of-practice into developing predictive models

Quick overview of other work

- Models for *personalized* prognosis in Complex, Chronic Conditions
- Physiologic models for predicting illness severity in infants
- Smart phone based illness severity monitoring

Where are the current gaps?

Where are the gaps?

1. **Methods that account for Treatments and Provider Practice Patterns**
    1. **Factored models over monolithic blackboxes**
    2. **Measure "regions of reliability" of the model**

2. **Methods for systematically integrating heterogeneous data**
    1. **Static with time-varying markers**
    2. **Handle different granularities at which data are sampled**
    3. **Data as they arrive over time**

3. **Personalize to the context in which they are being used: Practice-cost sensitive approaches**
    1. **When to act based on cost-benefit tradeoff**
    2. **Multiple models are deployed: utility based measures to alert**

4. **Methods for modeling heterogeneous populations**

"Variability is the law of life, and as no two faces are the same, so... no two individuals react alike and behave alike under the abnormal conditions which we know as disease."

- William Osler

Thank you!

ssaria AT cs.jhu.edu
twitter: @suchisaria