

# Deep Learning from Heterogeneous Sequences of Sparse Medical Data for Early Prediction of Sepsis

**Mahbub Ul Alam**<sup>1</sup>, Aron Henriksson<sup>1</sup>, John Karlsson Valik<sup>2, 3</sup>, Logan Ward<sup>4</sup>, Pontus Naucler<sup>2, 3</sup>, Hercules Dalianis<sup>1</sup>

<sup>1</sup>**Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden**

<sup>2</sup>Division of Infectious Disease, Department of Medicine, Karolinska Institute, Stockholm, Sweden

<sup>3</sup>Department of Infectious Diseases, Karolinska University Hospital, Stockholm, Sweden

<sup>4</sup>Treat Systems ApS, Aalborg, Denmark

**mahbub@dsv.su.se**

aronhen@dsv.su.se

john.karlsson-valik@sll.se

lw@treatsystems.com

pontus.naucler@ki.se

hercules@dsv.su.se



# Sepsis, At A Glance

- ❑ A life-threatening complication to infections
- ❑ One of the most serious forms of healthcare associated infections
- ❑ A leading cause of hospital morbidity and mortality
- ❑ Survival is dependent on initiating appropriate antimicrobial treatment as early as possible
- ❑ Mortality from septic shock increases by **7.6%** for every hour that antimicrobial treatment is delayed after the onset

# Goal

- ❑ Early prediction of sepsis in the non-ICU (intensive care units) setting from electronic health records (EHRs)
- ❑ Performance analysis of long short-term memory based recurrent neural network (RNN-LSTM)
- ❑ Investigating temporality and sequence length
- ❑ Investigating missingness

# HEALTH BANK - Swedish Health Record Research Bank

- ❑ Unique research resource containing a large sets of electronic patient records
- ❑ Used in a number of research projects carried out by the Clinical Text Mining Group, Department of Computer and Systems Sciences, Stockholm University
- ❑ Contains data from over 512 clinical units from Karolinska University Hospital (2006–2014) over two million patients.
- ❑ Structured information contains, a serial number (de-identified) for each patient, age, gender, ICD-10 diagnosis codes, drugs, ab and blood values, admission and discharge time, and date
- ❑ Unstructured data contains text written under different headings

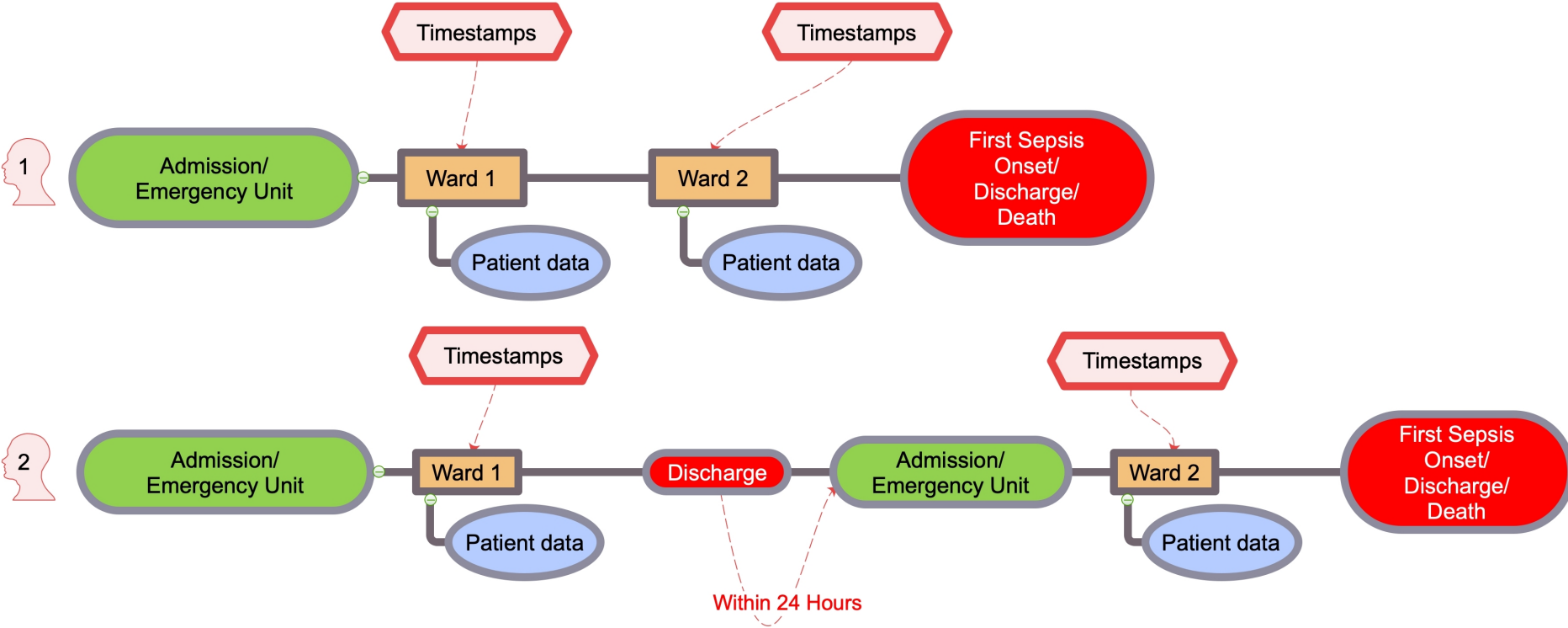


# Data

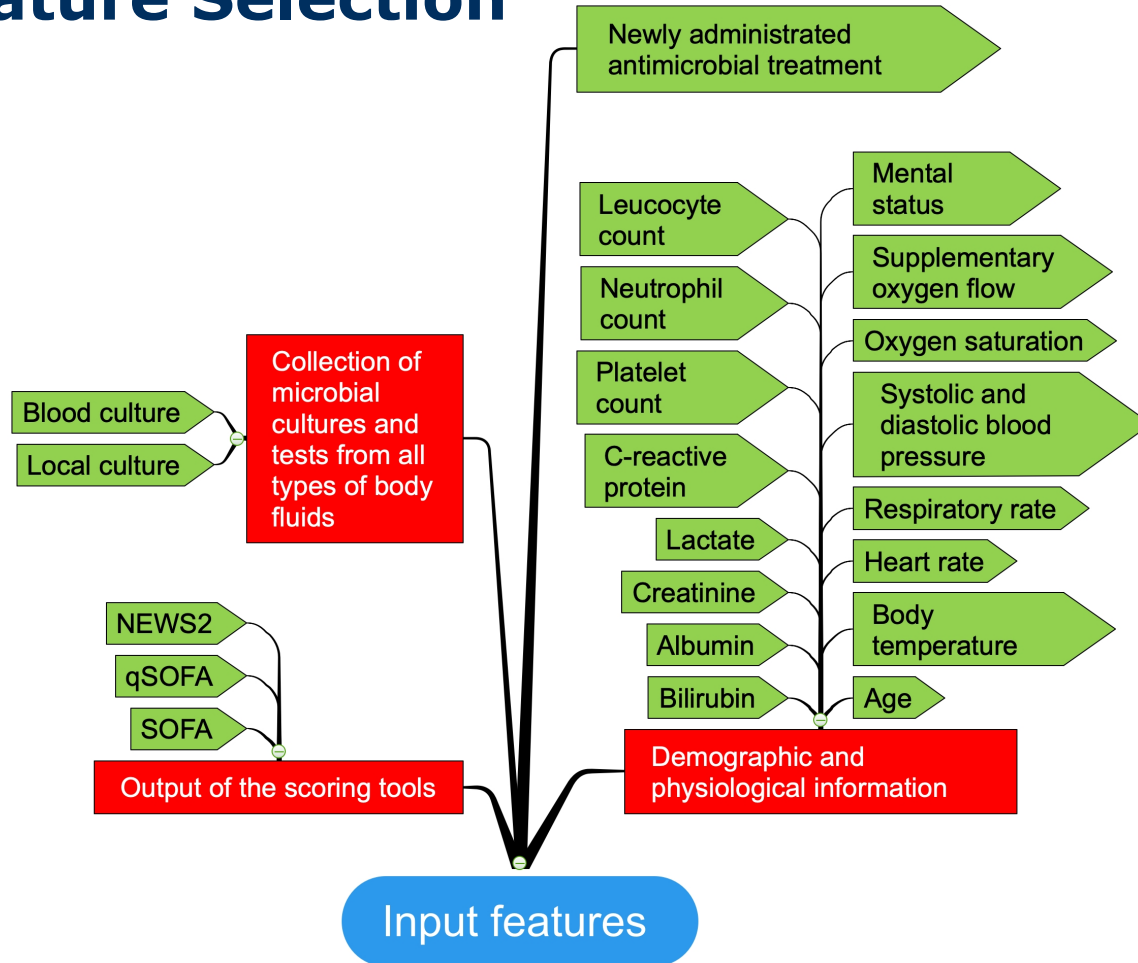
- ❑ Patients > 18 years admitted to the hospital between July 2010 and June 2013
- ❑ Followed until first sepsis onset, discharge or death
- ❑ Excluded if admitted to an obstetric ward and censored during ICU-care
- ❑ Encompasses **124,054 patients** and **198,638 care episodes**
- ❑ Sepsis in the cohort is **8.9%**



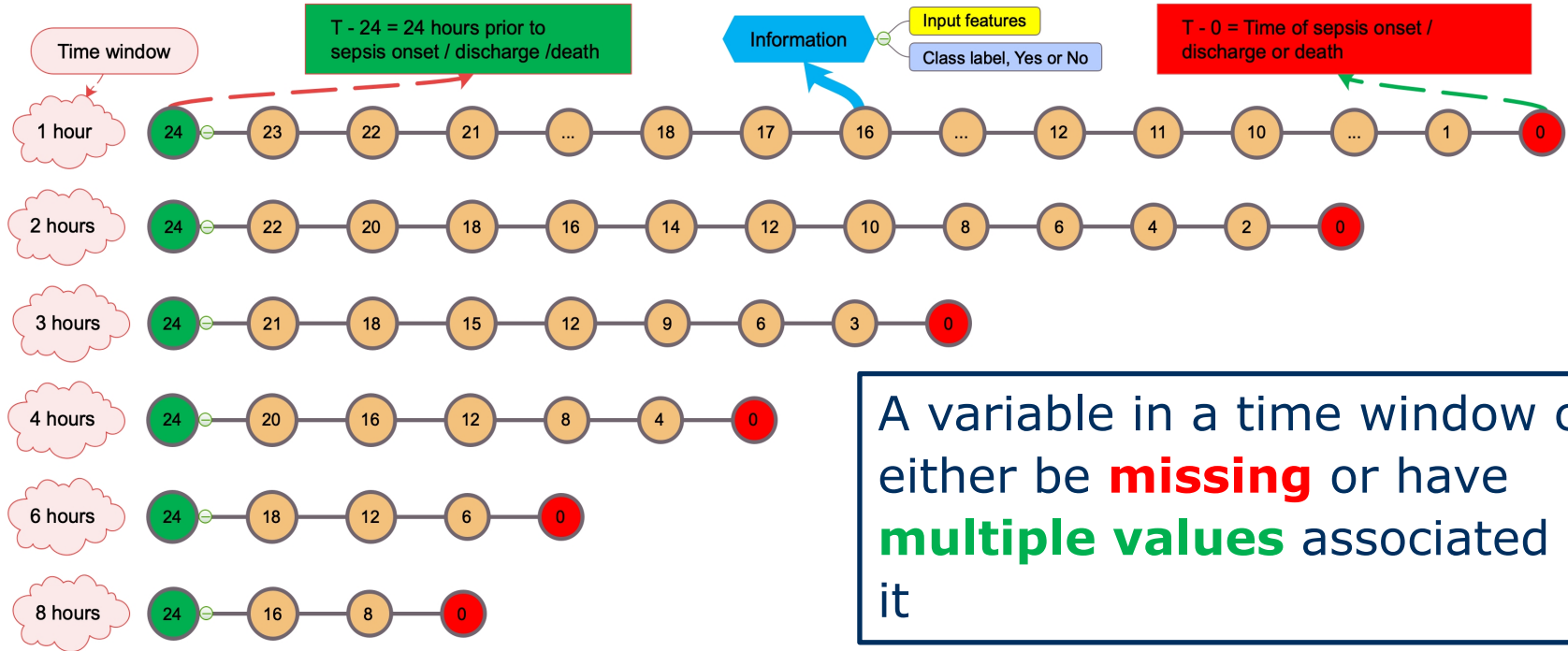
# Care Episode



# Input Feature Selection



# Care Episode Representation



When **multiple values** are present in a time window, the "**worst**" value is chosen



# Handling Missing Values, Without Imputation

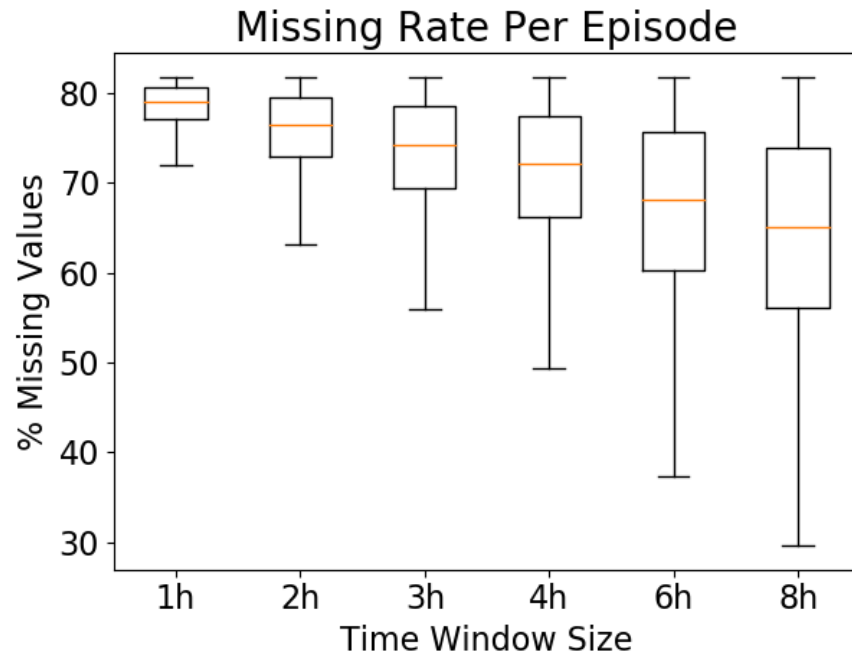
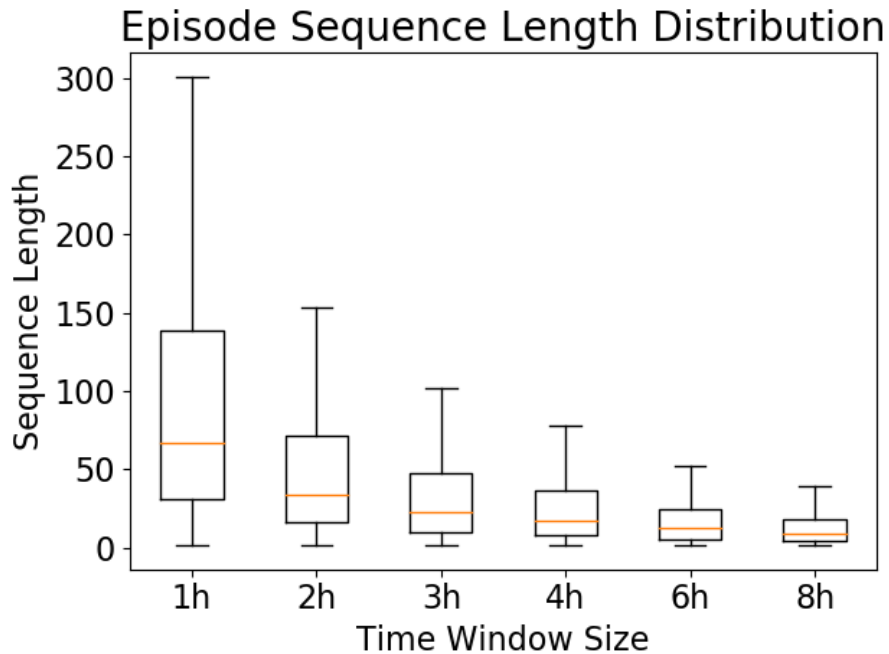
- When data is assumed to be **missing not at random**
  - ❖ Imputation is not carried out
  - ❖ Missing values are simply assigned an integer value which is not present in the data
  - ❖ The idea is that the model may learn to treat missingness as a distinct feature

# Handling Missing Values , With Imputation

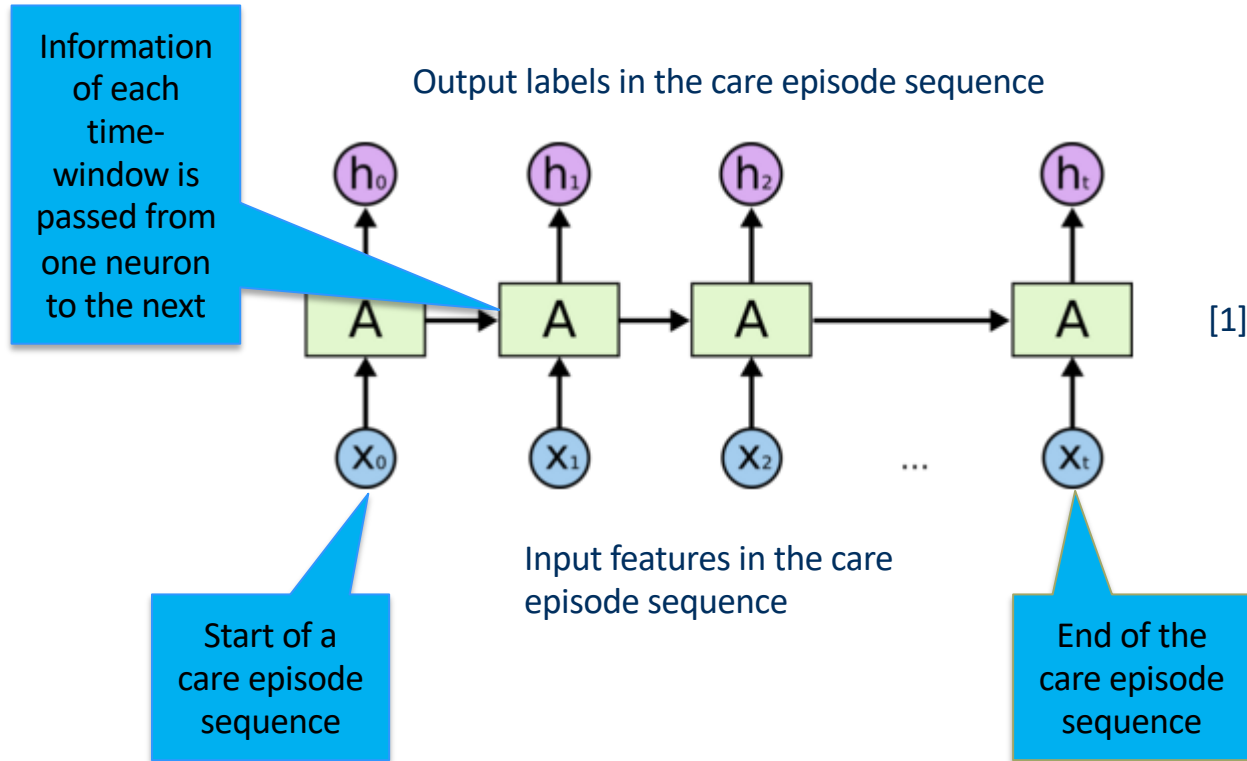
□ When data is assumed to be **missing at random**

- ❖ When a value exists for a given feature in the care episode, it is carried forward to subsequent windows until another present value is encountered, which is then in turn carried forward and so on
- ❖ When there is no value for a given feature in a care episode, it is imputed globally
  - For categorical features, the most frequent value is chosen, while mean imputation is carried out for numeric features
- ❖ For SOFA, qSOFA, and NEWS2, missing values are not mean-imputed; instead, the score is assumed to be 0 - if missing - at the start of an episode and then carried forward

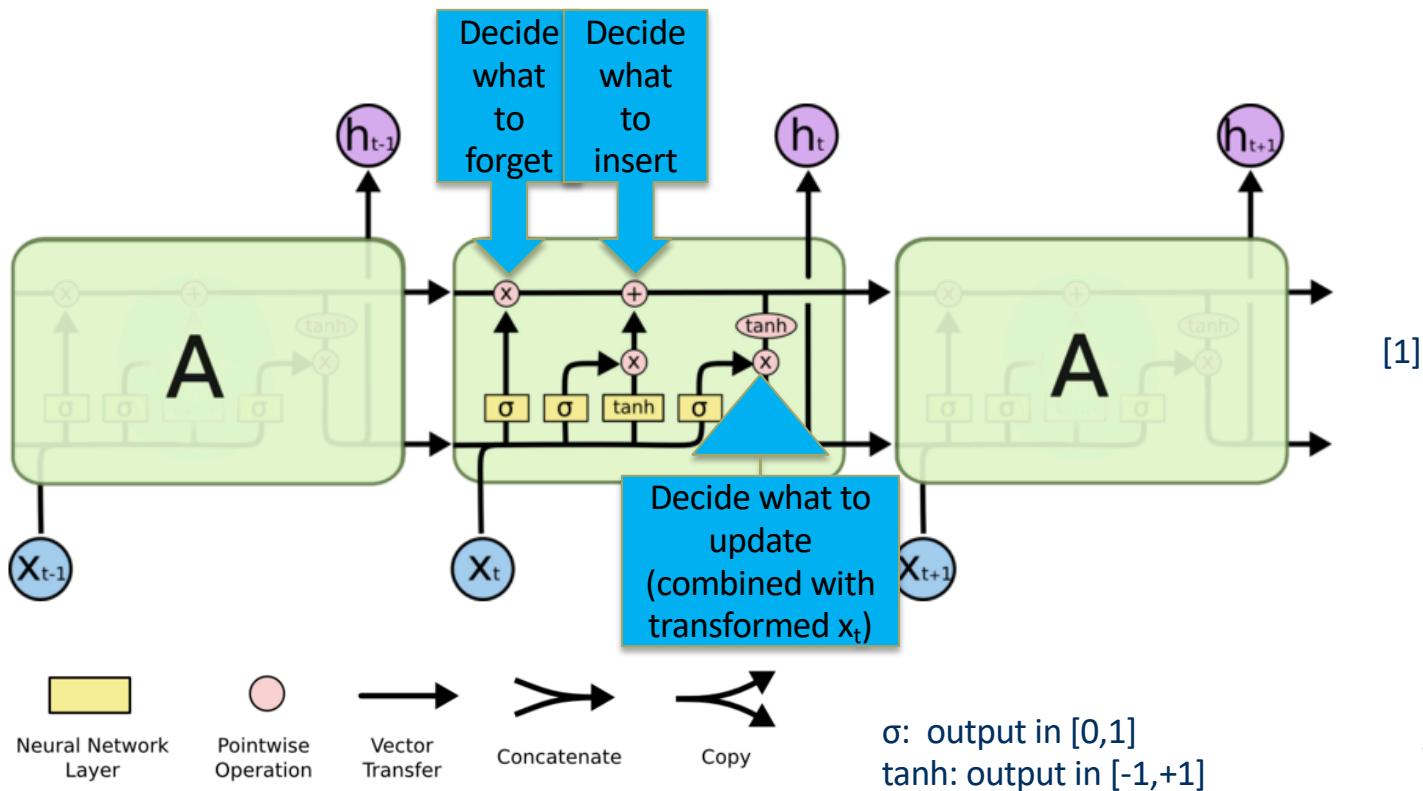
# Care Episode Distribution



# RNN (Recurrent Neural Network) Architecture



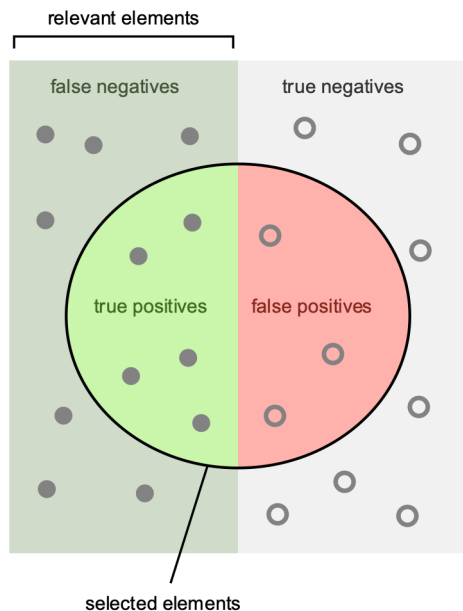
# LSTM (Long Short-Term Memory)-RNN Architecture



# Experiments

1. Different time window sizes
2. Handling missing values
3. Performance at different time points
4. Evaluation of earliness
5. Performance with different sequence lengths

# Evaluation Metrics: F<sub>1</sub>-Score



$$\textit{Precision} = \frac{TP}{TP + FP}$$

$$\textit{Recall (or Sensitivity or True Positive Rate)} = \frac{TP}{TP + FN}$$

$$\textit{Specificity} = \frac{TN}{TN + FP}$$

$$\textit{False Positive Rate (or } 1 - \textit{Specificity)} = \frac{FP}{TN + FP}$$

$$\textit{F}_1 \textit{ Score} = 2 \times \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}}$$

How many selected items are relevant?

Precision =



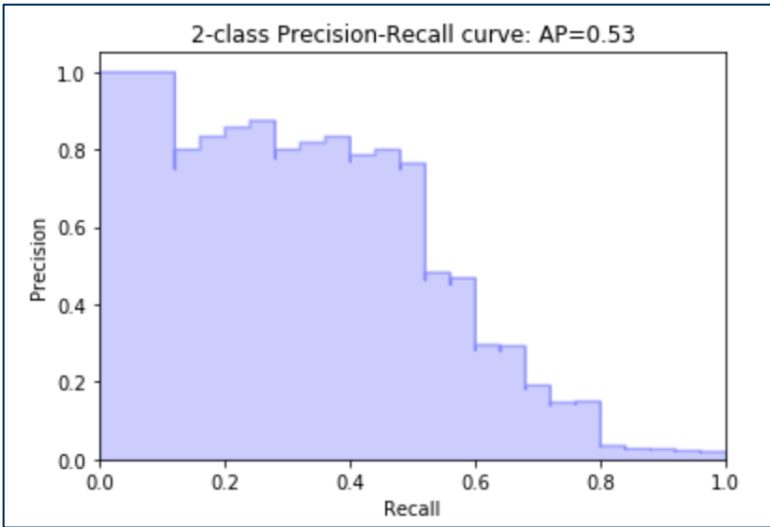
How many relevant items are selected?

Recall =



[2]

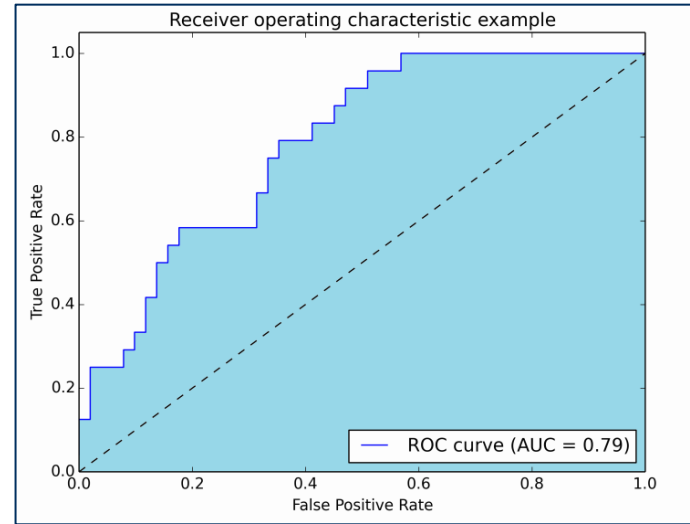
# Evaluation Metrics : AUPRC & AUROC



[3]

Area Under the Precision-Recall Curve (**AUPRC** or **AP**)-

Average of precision across all recall values



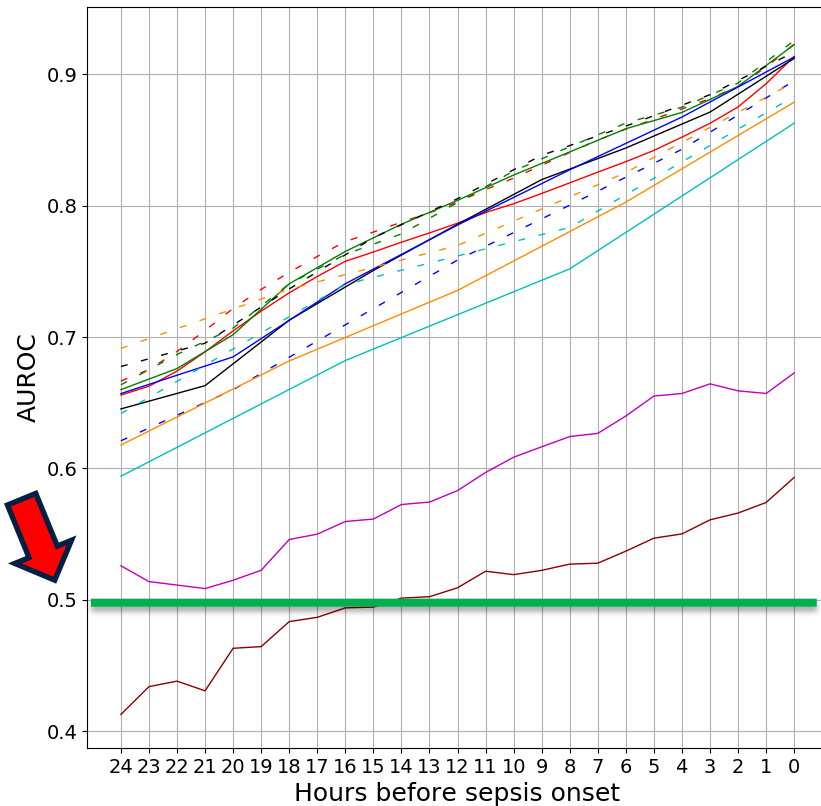
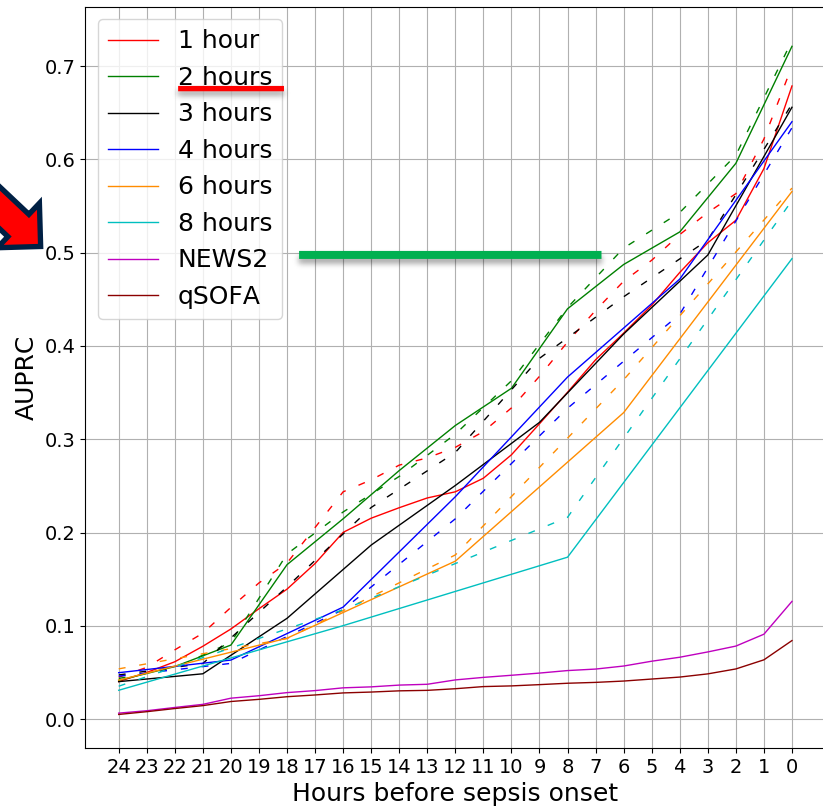
[4]

Area Under the Receiver Operating Characteristics (**AUROC** or **AUC**)-

Probability with which the classifier will assign a larger score to the positive than to the negative data point



# Predictive Performance At Different Time Points



--- without imputation  
— with imputation

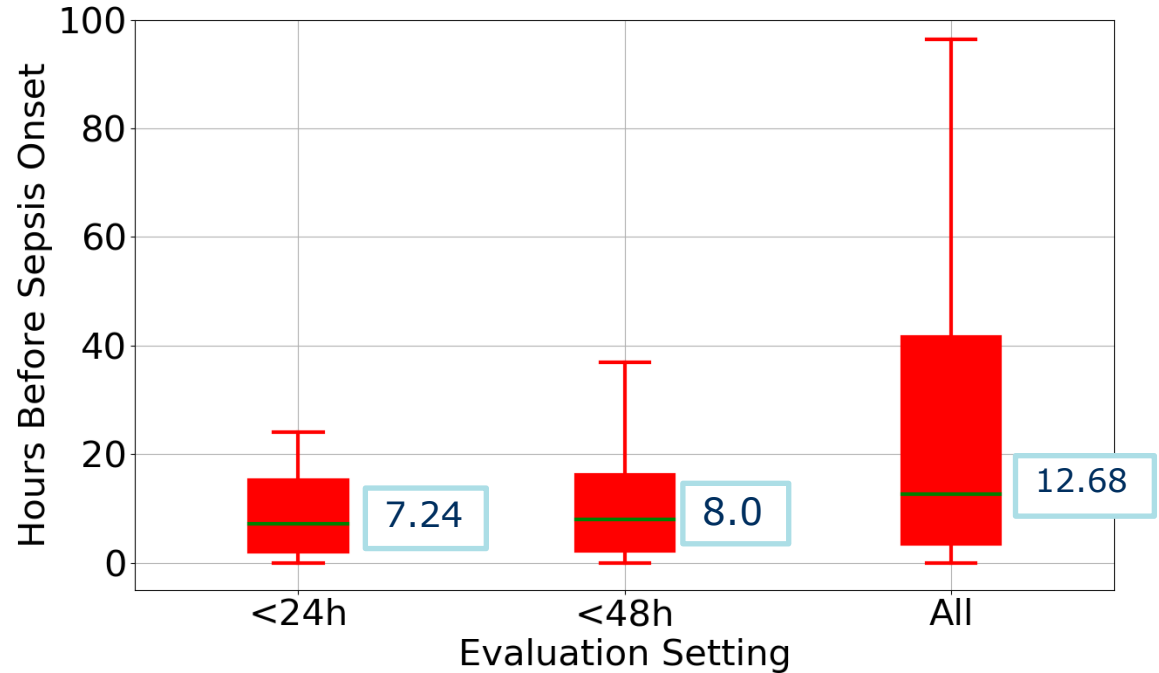
# Earliness Performance In Median Hours Before Sepsis Onset Combined With $F_1$ -score

Window Size	Without Imputation						With Imputation					
	<24h		<48h		All		<24h		<48h		All	
	Med.	$F_1$	Med.	$F_1$	Med.	$F_1$	Med.	$F_1$	Med.	$F_1$	Med.	$F_1$
1	6.00	0.65	7.37	0.53	11.79	0.28	5.90	0.61	7.77	0.48	10.78	0.25
2	7.24	0.73	8.00	0.57	12.68	0.33	7.02	0.74	8.00	0.57	13.43	0.32
3	12.45	0.77	12.57	0.65	14.34	0.38	7.78	0.75	8.33	0.65	13.00	0.42
4	7.13	0.73	7.92	0.63	14.22	0.41	11.22	0.79	10.00	0.67	13.36	0.41
6	10.09	0.76	11.00	0.68	14.00	0.46	8.43	0.74	8.48	0.65	14.00	0.44
8	8.73	0.74	9.82	0.67	14.75	0.46	6.58	0.70	8.00	0.64	24.28	0.44

Allowing true positives only **<24h** or **<48h** before sepsis onset, as well as **any time (All)**

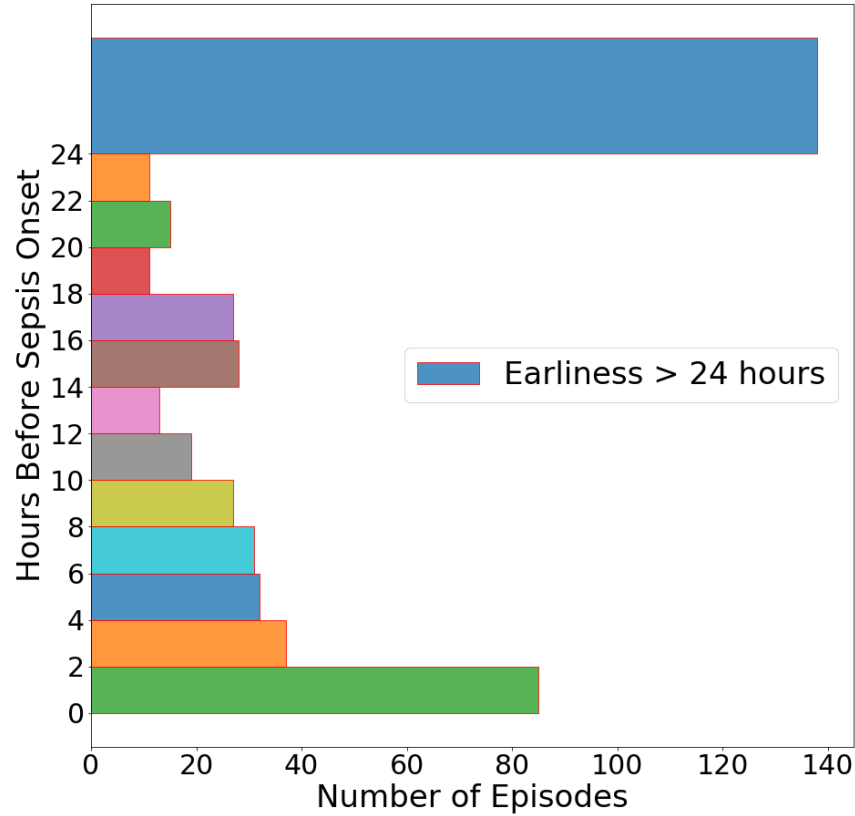
# Earliness Distribution

Two-hour time window and without imputation of missing values



# Episode-Wise Earliness

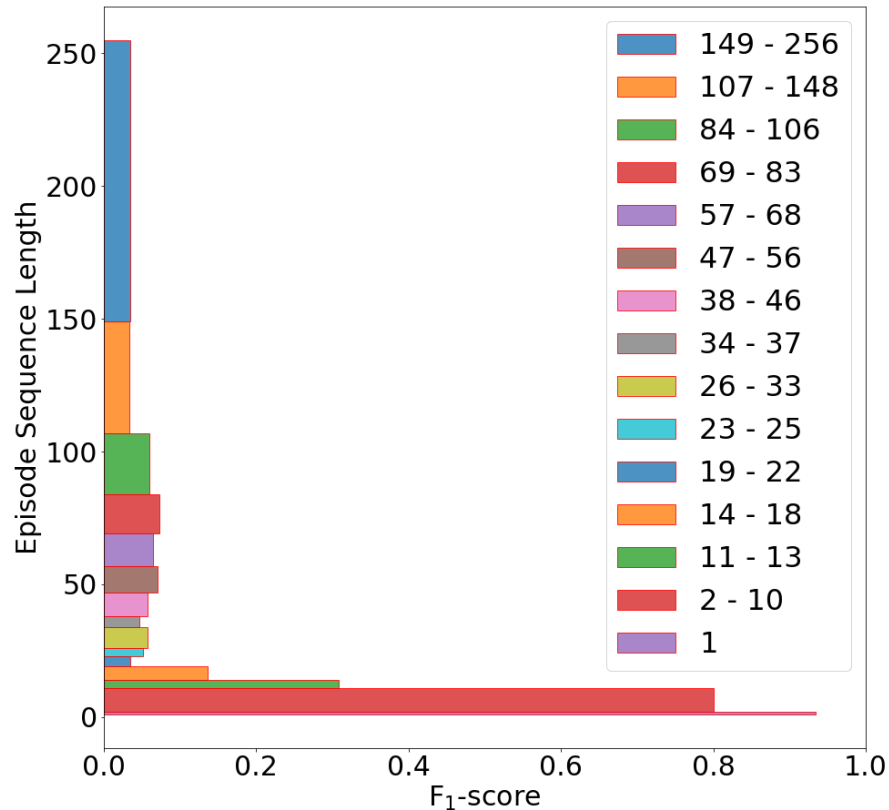
Two-hour time window and without imputation of missing values



# Length-Wise Performance Evaluation

Two-hour time window and without imputation of missing values

Each bin contains at least 1000 test instances



# Discussion

- ❑ Dividing the temporal EHR data into hourly time windows and prediction times
- ❑ Treating missing values in the care episodes as missing at random
- ❑ Early detection of sepsis in the non-ICU setting with sparse data
- ❑ Model performance on care episodes of different (sequence) lengths

# Future works

- ❑ Predicting sepsis should be divided into multiple stages to emulate the actual condition of sepsis
- ❑ We will investigate modifications to the neural architecture to make it more task-specific
- ❑ We will use additional natural language processing technique by incorporating free text data to provide additional important features to our model
- ❑ We also plan to investigate the interpretability of the models

# Questions?



# Image Source

[1] <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

[2] [https://en.wikipedia.org/wiki/Precision\\_and\\_recall](https://en.wikipedia.org/wiki/Precision_and_recall)

[3] <https://stackoverflow.com/questions/53772249/how-to-evaluate-accuracy-on-highly-unbalanced-data-using-naive-bayes-model>

[4] <https://riptutorial.com/machine-learning/example/14446/area-under-the-curve-of-the-receiver-operating-characteristic--auroc->

# Backup Slides



# Neural Network Parameters

Name	Values / Range
alpha	0, $10^{-4}$
beta one	0, $1 - 10^{-1}$
beta two	0, $1 - 10^{-3}$
hidden layers	2, 3, 4
neurons	64, 128, 256
drop out	0, 10, 20, 30, 40, 50, 60, 70
epochs	1, 2
mini-batch	100
classification function	log-softmax
optimizer	Adam optimizer

# NEWS2, National Early Warning Score 2

Chart 1: The NEWS scoring system

Physiological parameter	Score						
	3	2	1	0	1	2	3
Respiration rate (per minute)	≤8		9–11	12–20		21–24	≥25
SpO <sub>2</sub> Scale 1 (%)	≤91	92–93	94–95	≥96			
SpO <sub>2</sub> Scale 2 (%)	≤83	84–85	86–87	88–92 ≥93 on air	93–94 on oxygen	95–96 on oxygen	≥97 on oxygen
Air or oxygen?		Oxygen		Air			
Systolic blood pressure (mmHg)	≤90	91–100	101–110	111–219			≥220
Pulse (per minute)	≤40		41–50	51–90	91–110	111–130	≥131
Consciousness				Alert			CVPU
Temperature (°C)	≤35.0		35.1–36.0	36.1–38.0	38.1–39.0	≥39.1	

<https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2>

# NEWS2, National Early Warning Score 2

Chart 2: NEWS thresholds and triggers

NEW score	Clinical risk	Response
Aggregate score 0–4	Low	Ward-based response
Red score Score of 3 in any individual parameter	Low–medium	Urgent ward-based response*
Aggregate score 5–6	Medium	Key threshold for urgent response*
Aggregate score 7 or more	High	Urgent or emergency response**

\* Response by a clinician or team with competence in the assessment and treatment of acutely ill patients and in recognising when the escalation of care to a critical care team is appropriate.

\*\*The response team must also include staff with critical care skills, including airway management.

<https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2>

# Sequential Organ Failure Assessment (SOFA) score

## The SOFA Score\*

Organ System, Measurement	SOFA Score				
	0	1	2	3	4
Respiration PaO <sub>2</sub> /FiO <sub>2</sub> , mmHg	Normal	<400	<300	<200 (with respiratory support)	<100 (with respiratory support)
Coagulation Platelets x10 <sup>3</sup> /mm <sup>3</sup>	Normal	<150	<100	<50	<20
Liver Bilirubin, mg/dL (μmol/l)	Normal	1.2-1.9 (20-32)	2.0-5.9 (33-101)	6.0-11.9 (102-204)	>12.0 (<204)
Cardiovascular Hypotension	Normal	MAP<70 mmHg	Dopamine ≤5 or dobutamine (any dose)**	Dopamine >5 or epinephrine ≤0.1 or norepinephrine ≤0.1	Dopamine >15 or epinephrine >0.1 or norepinephrine >0.1
Central Nervous System Glasgow Coma Score	Normal	13-14	10-12	6-9	<6
Renal Creatinine, mg/dL (μmol/l) or Urine output	Normal	1.2-1.9 (110-170)	2.0-3.4 (171-299)	3.5-4.9 (300-440) or <500 mL/day	>5.0 (>440) or <200 mL/day

\* Source: Vincent et al., 1996.

\*\*Adrenergic agents administered for at least 1 hour (doses given are in mcg/kg/min).

[http://www.iccueducation.org.uk/uploads/2/3/1/0/23109338/sofa\\_score.pdf](http://www.iccueducation.org.uk/uploads/2/3/1/0/23109338/sofa_score.pdf)



# quick Sepsis related Organ Failure Assessment (qSOFA)

Assessment	qSOFA score
Respiratory rate $\geq 22$ /min	1
Altered mentation*	1
Systolic blood pressure $\leq 100$ mm Hg	1

\*Glasgow Coma Scale (GCS) $\leq 14$

<https://jamanetwork.com/journals/jama/fullarticle/2492881>

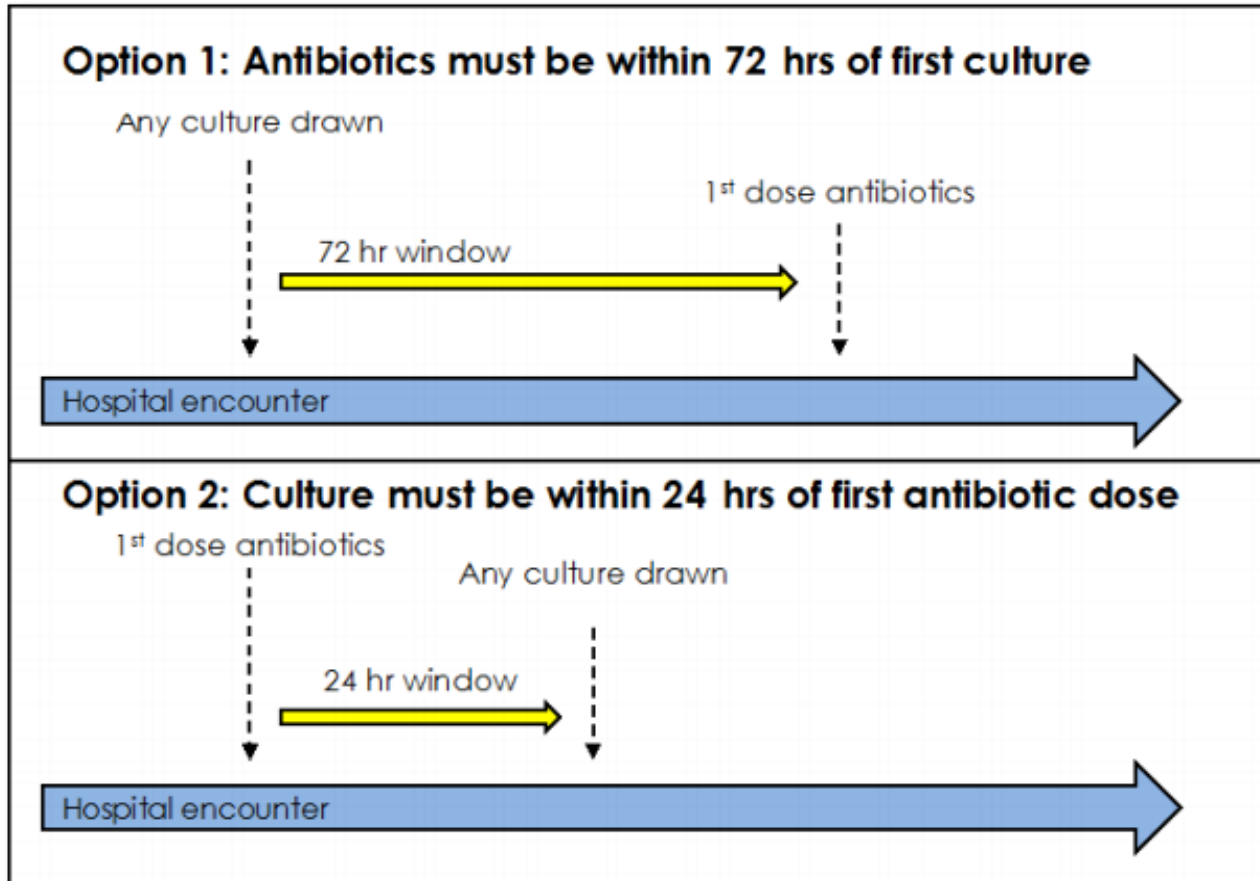
<https://www.glasgowcomascale.org/>

# Sepsis Definition, Sepsis-3 clinical criteria

- ❑ **Suspected infection** in combination with **organ dysfunction**
- ❑ **Sepsis onset time** is the first time window when both ***organ dysfunction*** and ***suspected infection*** criteria are met (time zero)



# Sepsis Definition, Suspected Infection



# Sepsis Definition, Organ Dysfunction

- ❑ **SOFA score**  $\geq 2$  compared to the baseline
- ❑ Measured 48 hours before to 24 hours after the onset of **suspected infection**
- ❑ Baseline SOFA score is defined as the latest value measured before the 72-hour time window and is assumed to be 0 in patients not known to have a pre-existing organ dysfunction