MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Lost in translation: Challenges in transferring global state-of-the-art AI models to Australian hospitals

Stephanie Chaousis, PhD

Chief Customer Officer

Datarwe

Stephanie Chaousis

@steph_chaousis

Datarwe.com

@TheInstituteDH #MEDINFO23

AIDH IMIA

Datarwe's Clinical Data Nexus

The Clinical Data Nexus (CDN) is a trusted, secure data utility that allows accelerated connection of multi-modal data sets for approved users.

1 PLATFORM

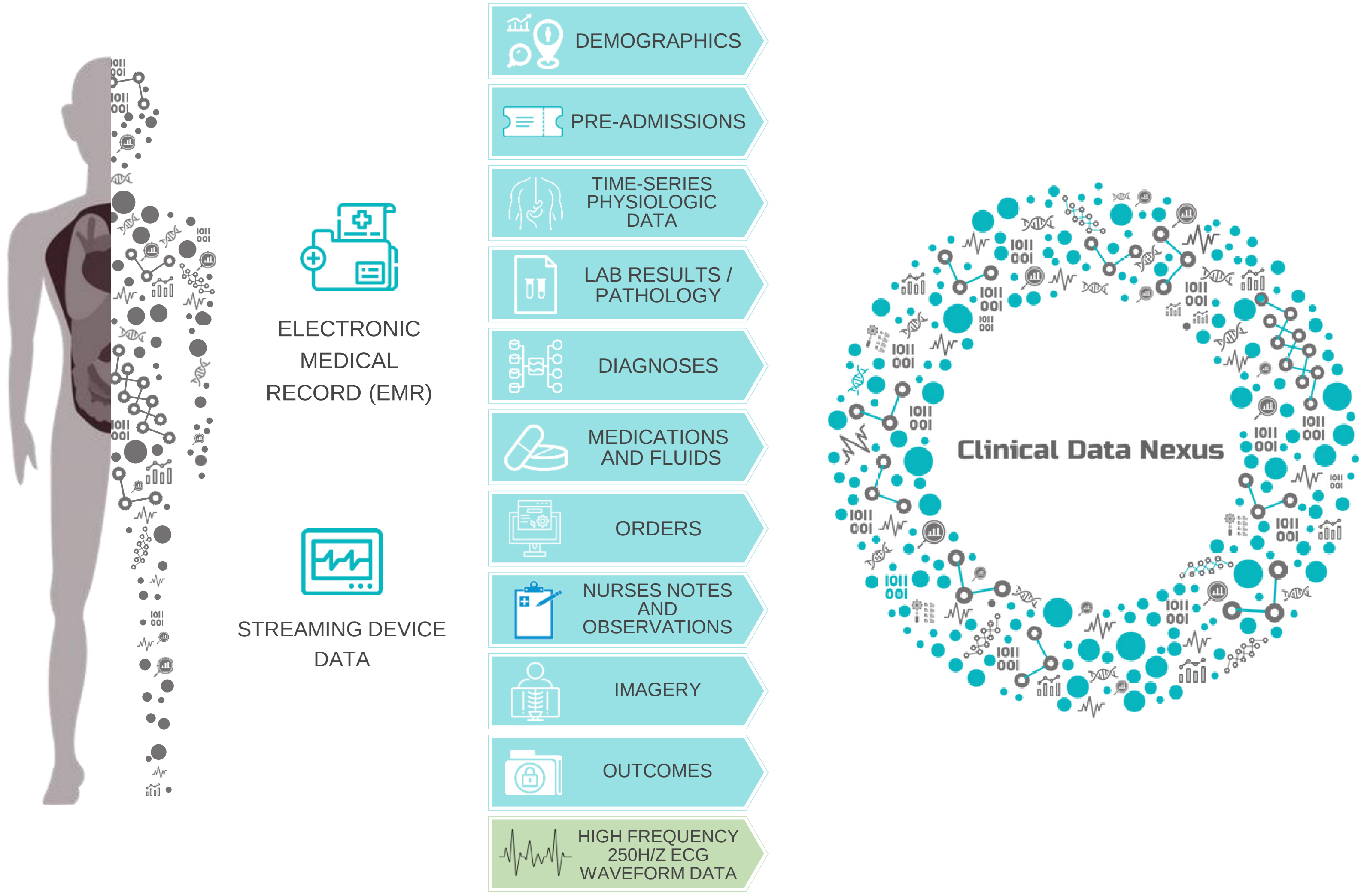
Datarwe's CDN is a singular point of connection to a vast network of local and international real world data

19+ HOSPITALS

Datarwe's CDN links 19+ national and international hospitals

13M+ EPISODE HRS

Datarwe's CDN avails network participants to more than 13 million episode hours of curated data



The promise of AI for healthcare

- Democratisation of healthcare access
- Improved patient outcomes
- Reduced clinician workload
- Increased capacity for quality service delivery

Predictive models for clinical decision support

Sepsis

#1 cause of hospital deaths

$38B cost to healthcare annually

Article | Published: 21 July 2022

nature medicine

Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis

Roy Adams, Katharine E. Henry, Anirudh Sridharan, Hossein Soleimani, Andong Zhan, Nishi Rawat, Lauren

Reduced mortality

14% vs 19%

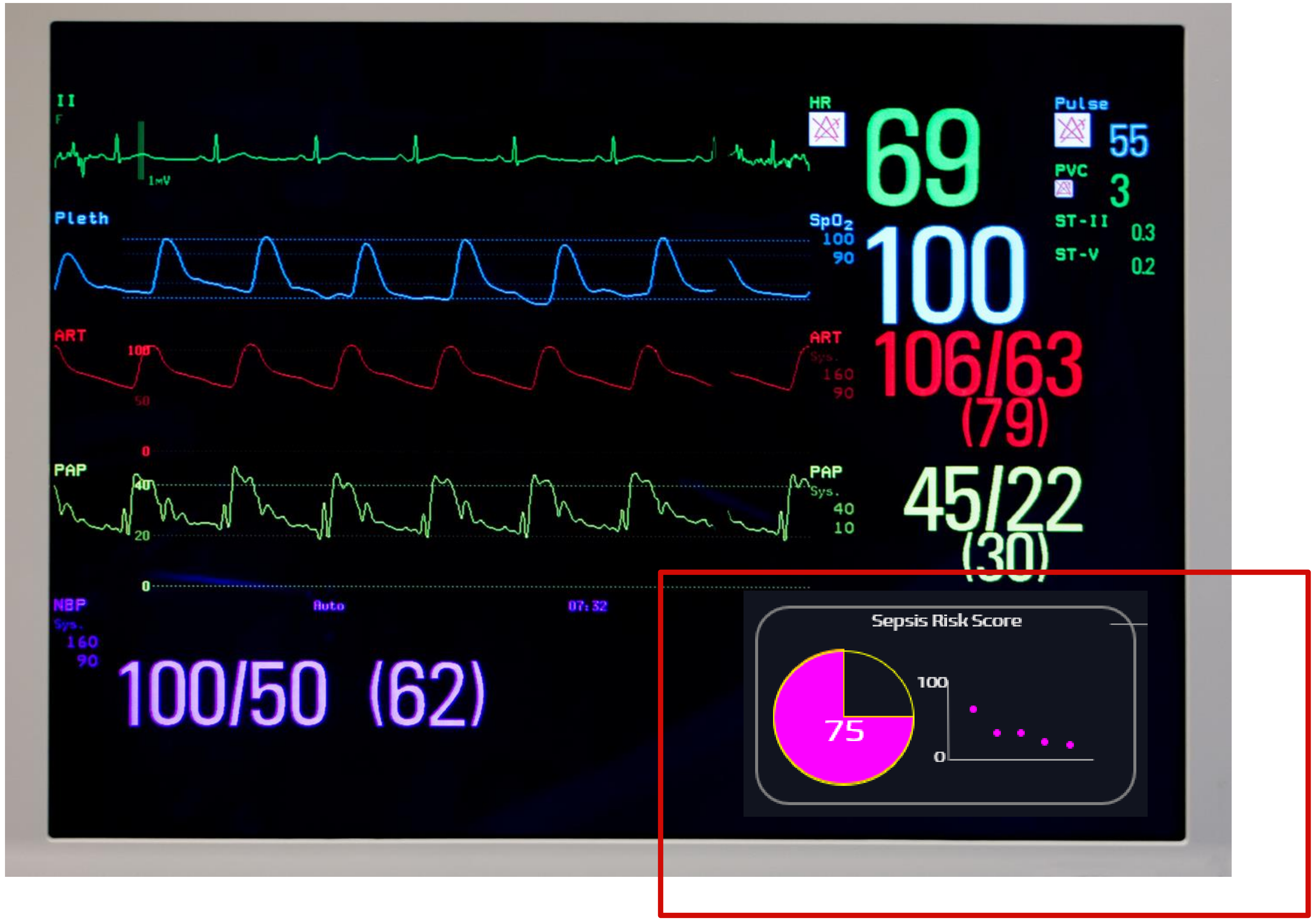
Improved SOFA score progression

−0.8 vs −0.4

Reduced length of stay

6.6d vs 8.1d

Use case: personalised live sepsis risk monitoring



datarwe

Deploying state-of-the-art models: why re-invent the wheel?

Approach

Replicate study results on original data set (MIMIC-III)

Attempt to reproduce early detection of sepsis accuracy (MIMIC-III)

Test transferibility on retrospective clinical data (AUS ICU)

Outcomes

- Could not replicate reported AUC from study on original data.
- Transferability failed despite extensive testing.

JAMA Internal Medicine

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients

Andrew Wong, MD[1]; Erkin Otles, MEng[2,3]; John P. Donnelly, PhD[4]; et al

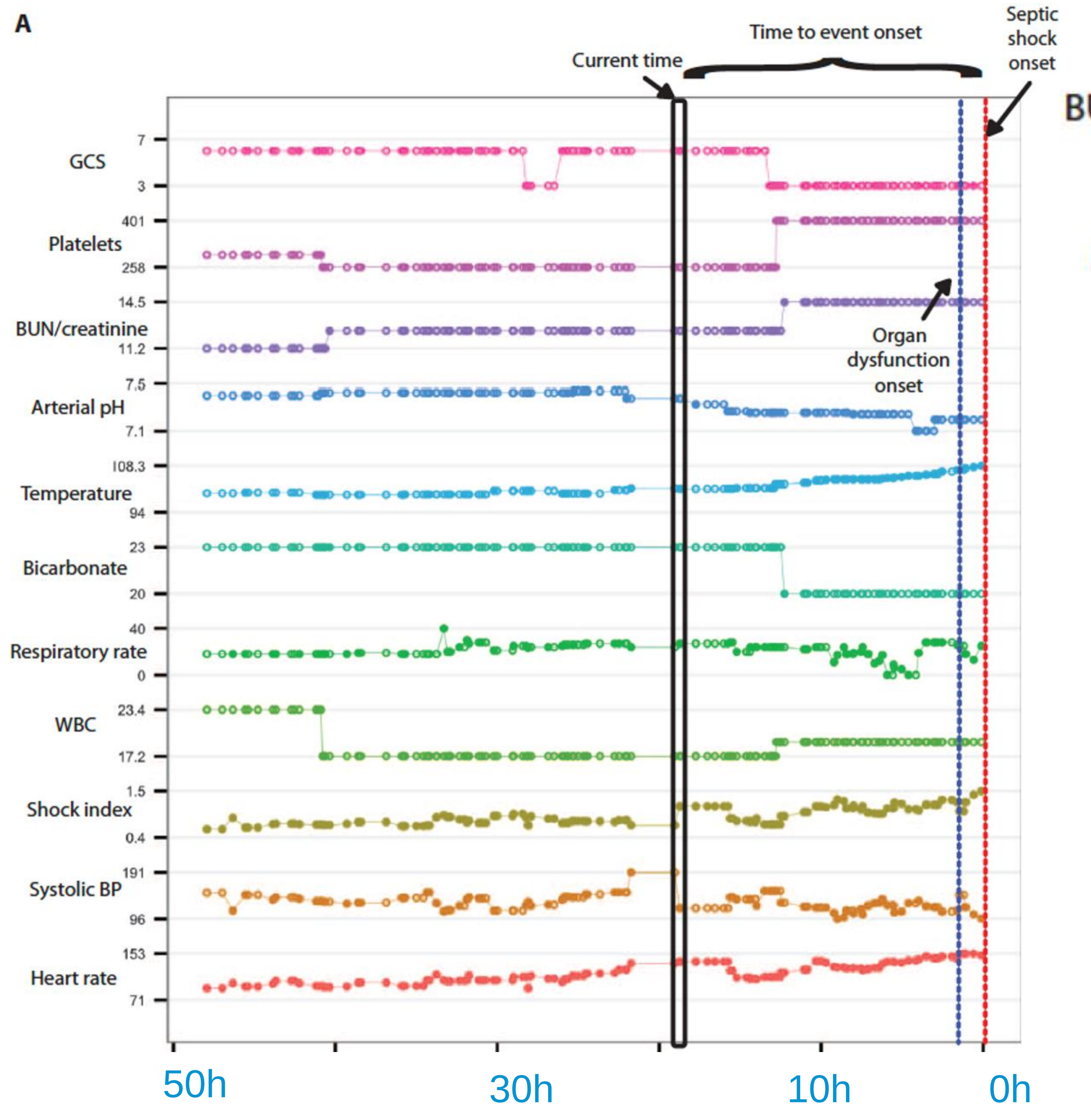
- Only identified 183 of 2552 patients with sepsis (7%) who did not receive timely treatment
- Low sensitivity in comparison with clinical practice.
- Did not identify 1709 patients with sepsis (67%)

MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Parameters used to calculate risk score

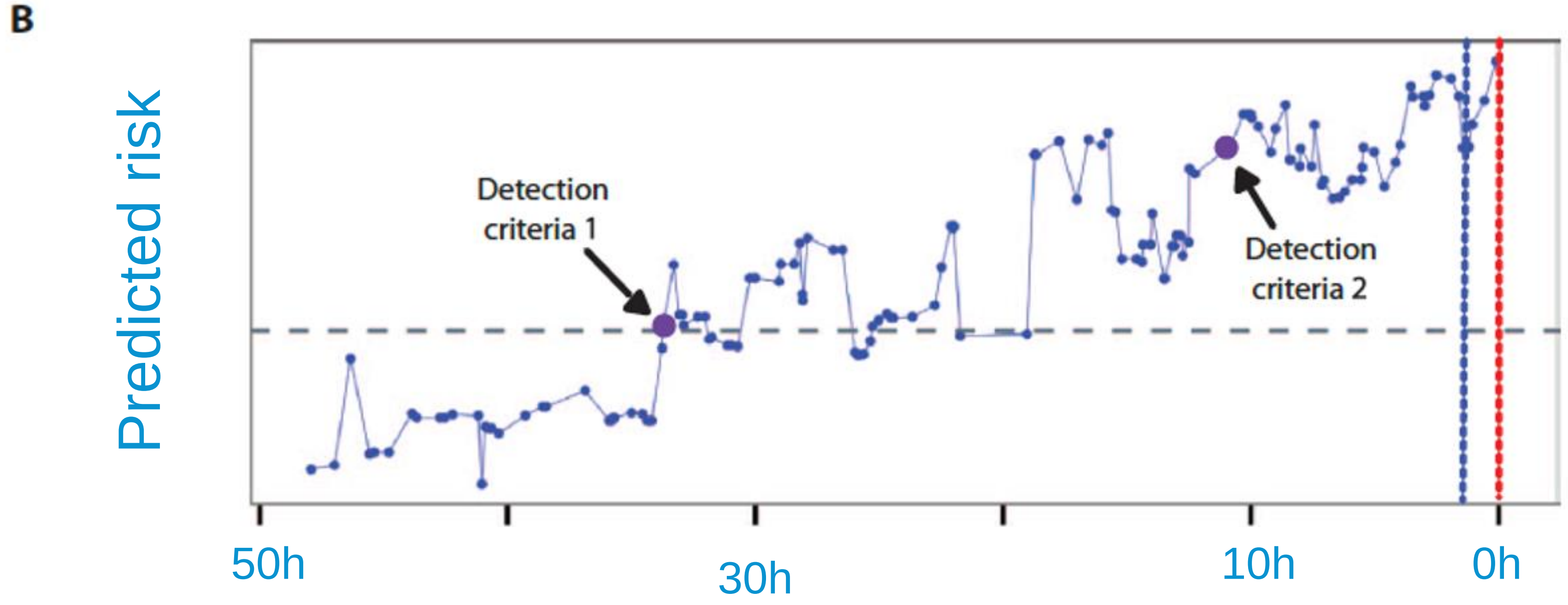
Input values from EMR



Hours until septic shock onset

How long is the patient's risk score maintained above a certain threshold?

Predicted risk



Hours until septic shock onset

eDH

datarwe

Model translation: challenges at each stage

REPRODUCIBILITY	INPUT DATA	VALIDATION	CALIBRATION	PROSPECTIVE TRIAL	PRODUCTION MONITORING
• Can the results be reproduced on the original data set?	• Can the model be applied to the target data set?	• Can the results be validated on real-world-data of the target site?	• Can calibration be achieved without losing accuracy?	• Is the model robust enough to perform consistently in a real-world setting?	• How do we monitor the model after deployment?

Reproducibility

REPRODUCIBILITY

Can the results be reproduced on the original data set?

- How has the training data set been cleansed?
- How has the test data set been cleansed?
- What were the inclusion & exclusion criteria?
- What hyperparameters were used? (ML parameters)
- How are the models weighted?

Database | Credentialed Access

MIMIC-III Clinical Database

Alistair Johnson, Tom Pollard, Roger Mark

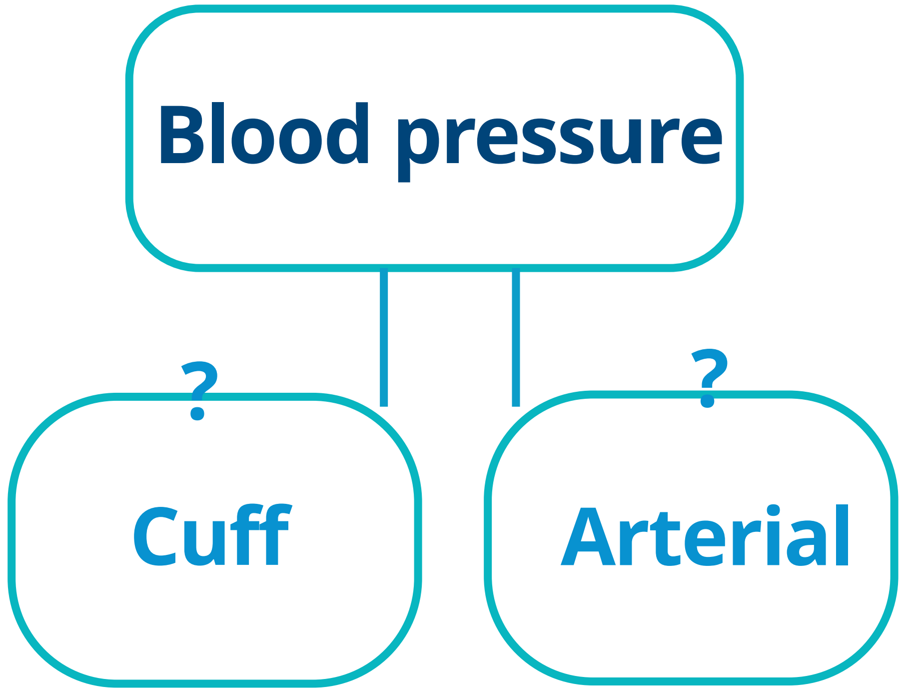
Published: Sept. 4, 2016. Version: 1.4

Input data

INPUT DATA

What data should be used for model testing?

- Variance in data across populations.
- Lack of model development transparency.
- Missing data variations.
- Data frequency and aggregating data.
- Challenges in standardizing and integrating. diverse data sources.



Validation

VALIDATION

Can the results be validated on real-world-data of the target site?

- How has the data been verified/ground truth labelling?
- How is the disease defined at the target site?
- Risk of overfitting to training data.
- Difficulty in defining suitable validation metrics eg. use of billing codes (ICD9).

Calibration

CALIBRATION

Can calibration be achieved without losing accuracy?

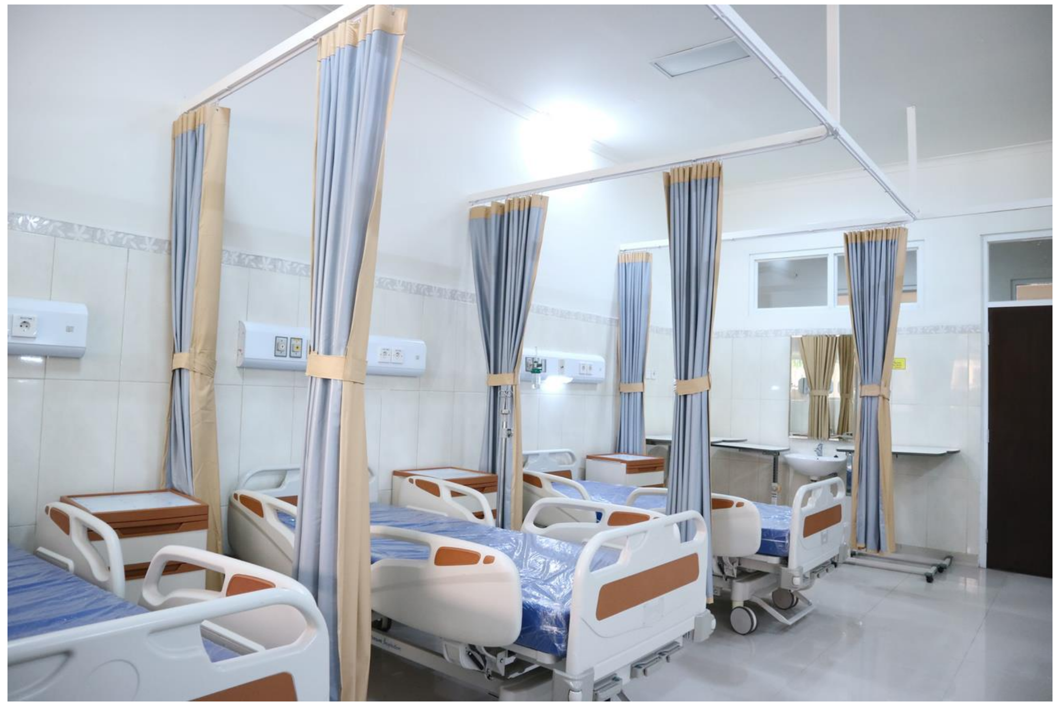
- Ensuring predicted probabilities align with real-world outcomes.
- Generalisation versus fine-tuned.
- Federated learning or transfer learning.
- Training data extraction.
- Is the volume of data required for accurate calibration available?

Prospective trial

PROSPECTIVE TRIAL

Is model robust enough to perform consistently in real-world setting?

- Time and cost-intensive real-world validation.
- Challenges in ethical trial design/control group.
- Ensuring patient safety in trial execution.
- Integration into human-in-the-loop.
- Bias: automation, selection.



Production monitoring

PRODUCTION MONITORING

How do we monitor the model after deployment?

- Model performance monitoring over time can be complex.
- Need for continuous updates and user feedback integration.
- Difficulties in integrating models into existing workflows and IT systems.

Conclusion

- **Translation of clinical AI tools is complex.**
- **A difficult problem to solve but critical for maintaining a modern healthcare system.**
- **Collaboration is key: clinicians & technologists.**
- **Access to aggregated, clean, labelled, real-world data is fundamental to success (Datarwe CDN).**

Datarwe's CDN provides access to data for **clinical model calibration & validation**

datarwe

Stephanie Chaousis, PhD

Chief Customer Officer

steph.chaousis@datarwe.com

Stephanie Chaousis

datarwe.com