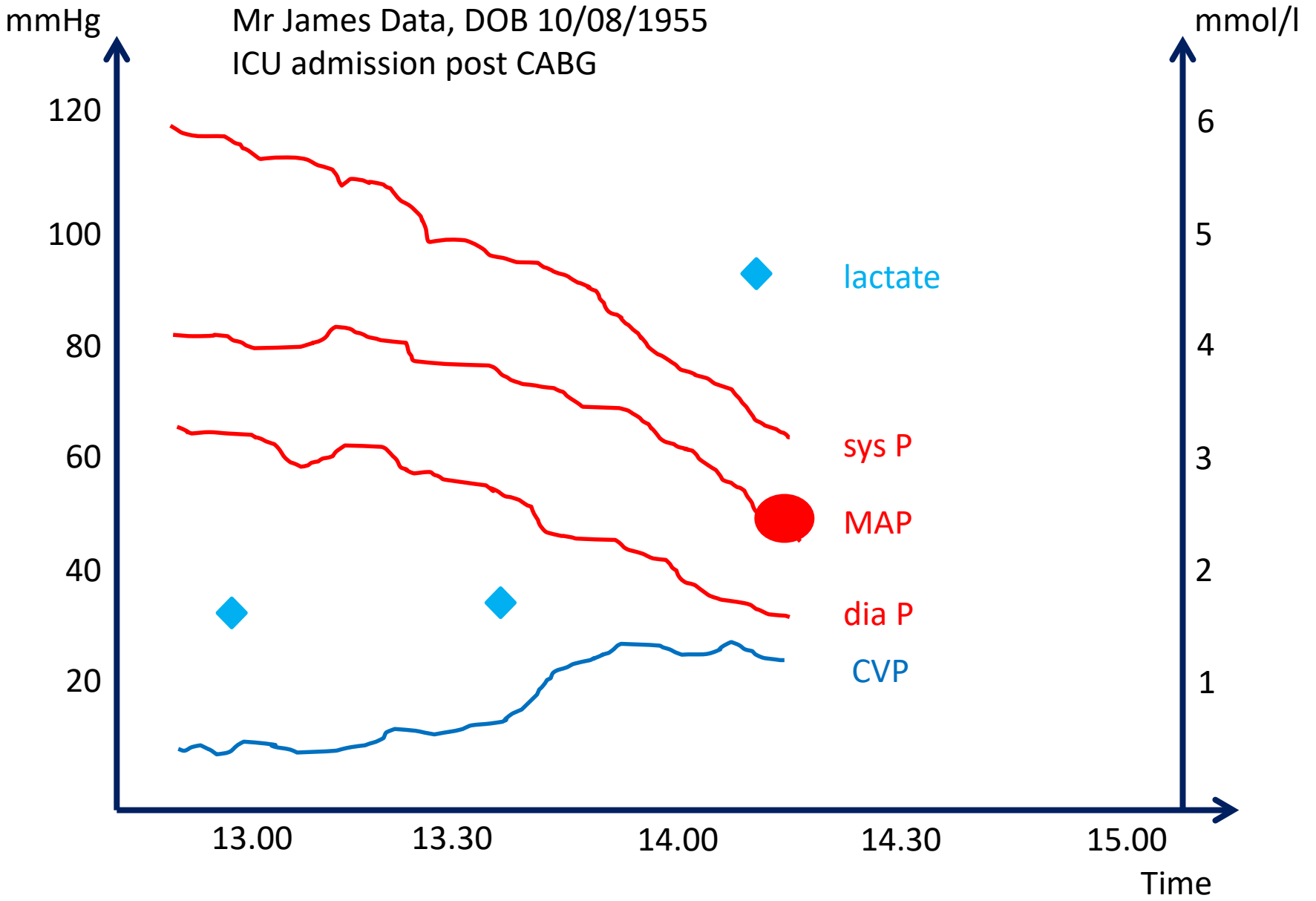
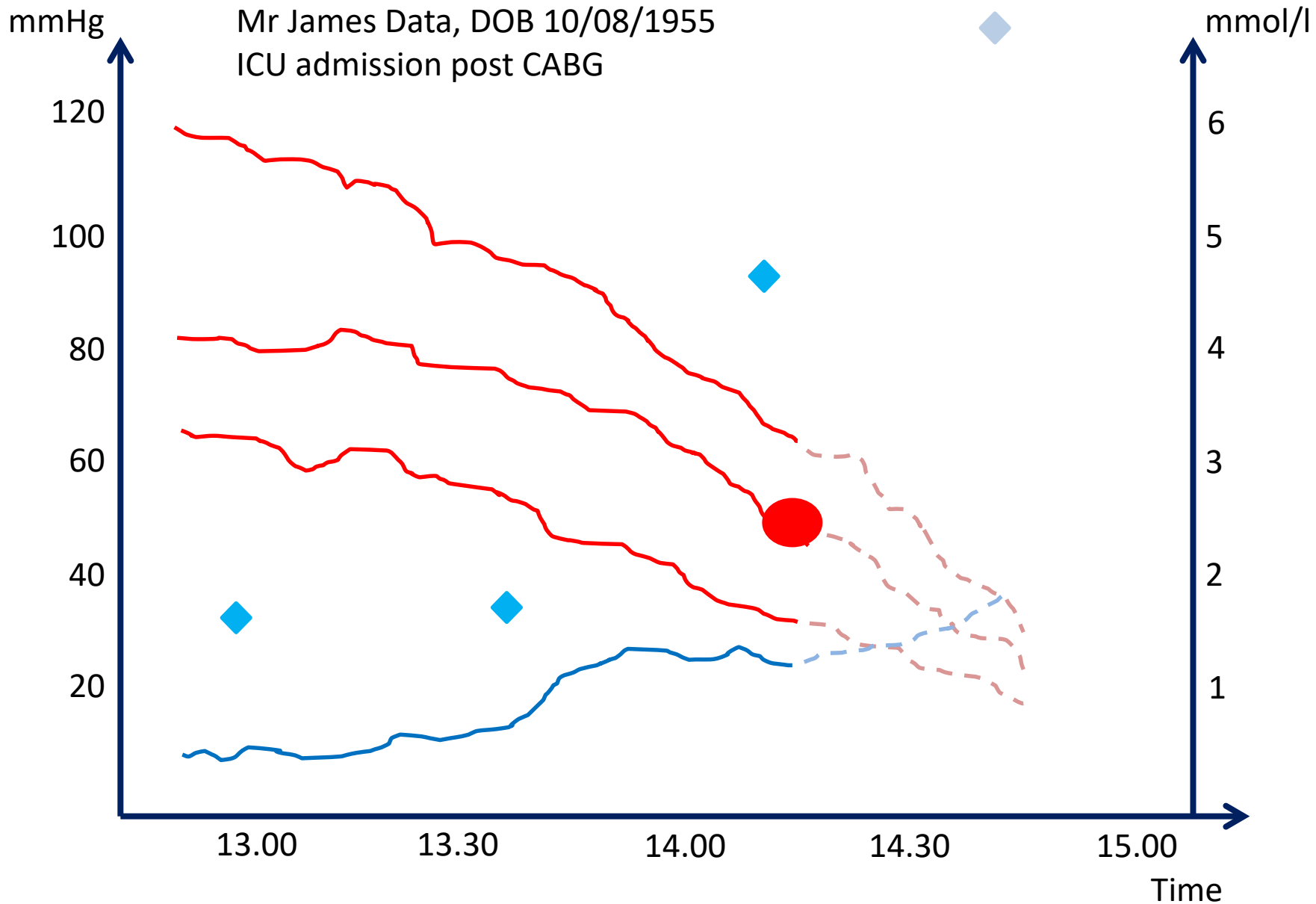


Building a High Resolution Clinical Database

Tobias Merz (ADHB), Alex Kazemi
(CMDHB), Shawn Sturland (MIT), Leo
Anthony Celi (MIT)





mmHg - mmol/l



prediction window

now

event

Big Data in ICU

- Critical illness = presence or risk of development of organ dysfunction/failure
- Organ functions are continuously monitored in a ICU
- Regular patient evaluations are important to assess for changes in organ function parameters in context of established treatments
- ICUs use electronic patient data management systems (PDMS) to store patient-specific data in a single database.
- The ICU is a data rich environment. Monitoring all changes over time in many patients is difficult.

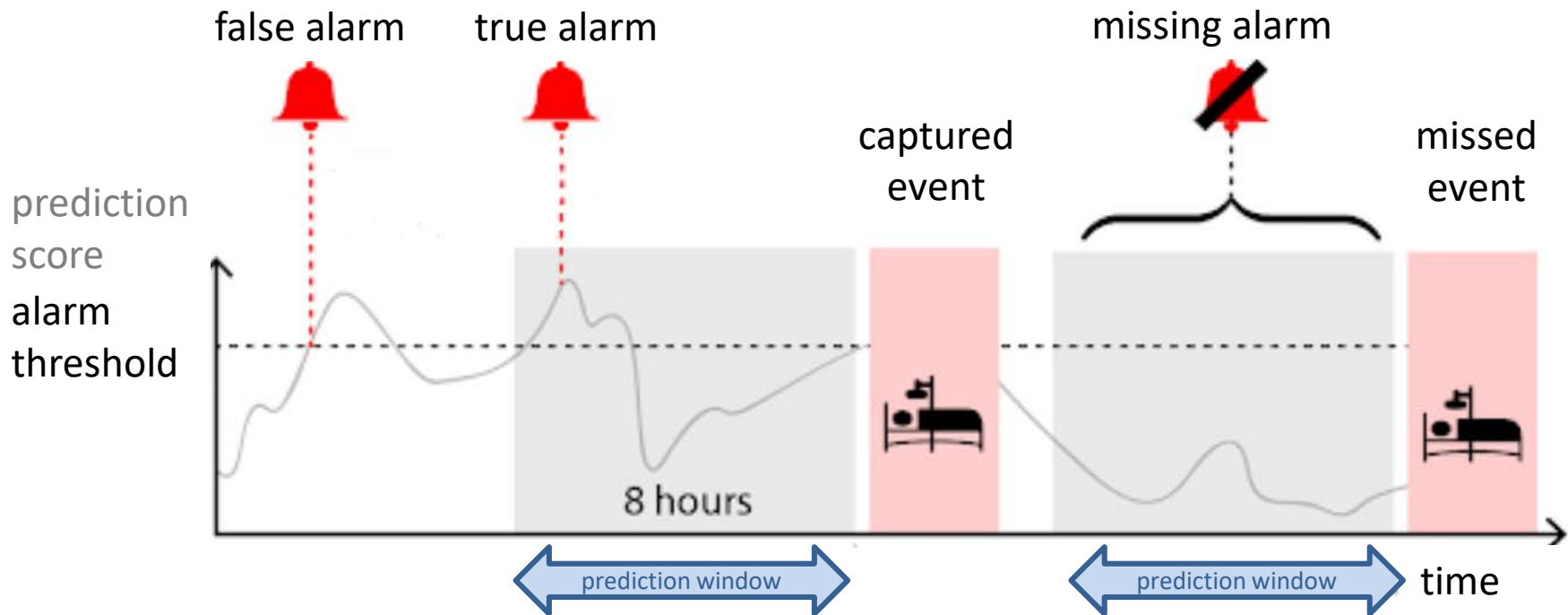
Example



Contents PDMS (Patient Data Management System)

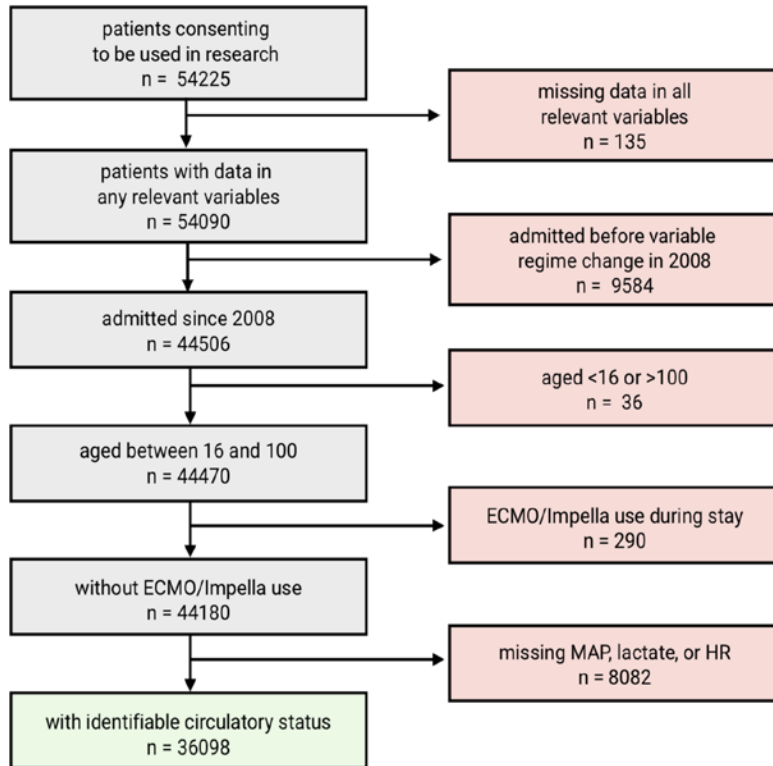
- Starting point:
 - Data collected between 04/2006 – 10/2016
 - 54'225 patient admissions
 - 7333 different variables
 - 3'124'435'321 data points
- A lot of noise...

Task and evaluation

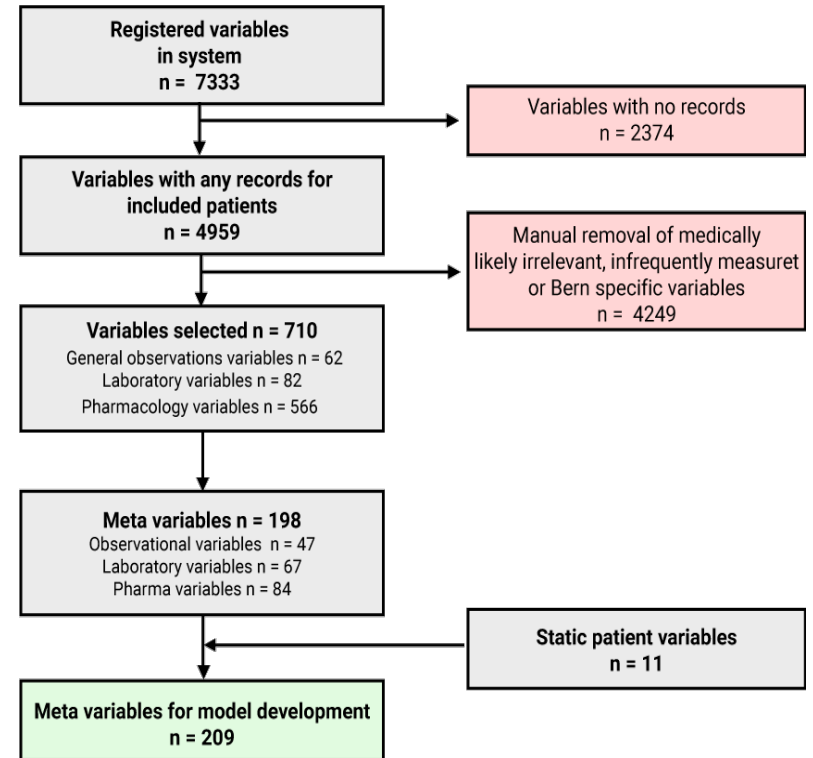


Data selection

Patient selection

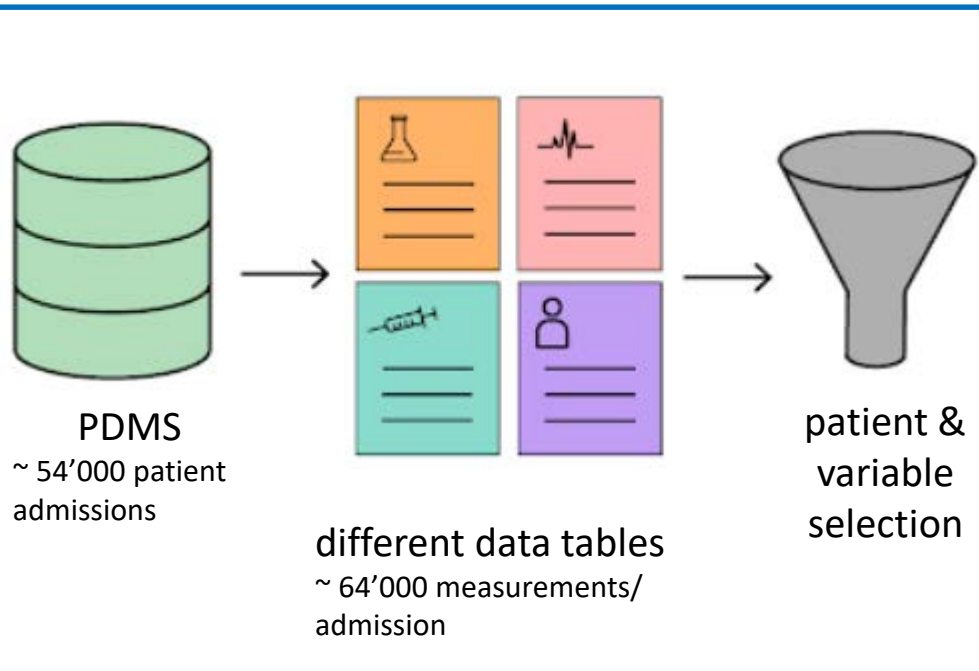


Variable selection

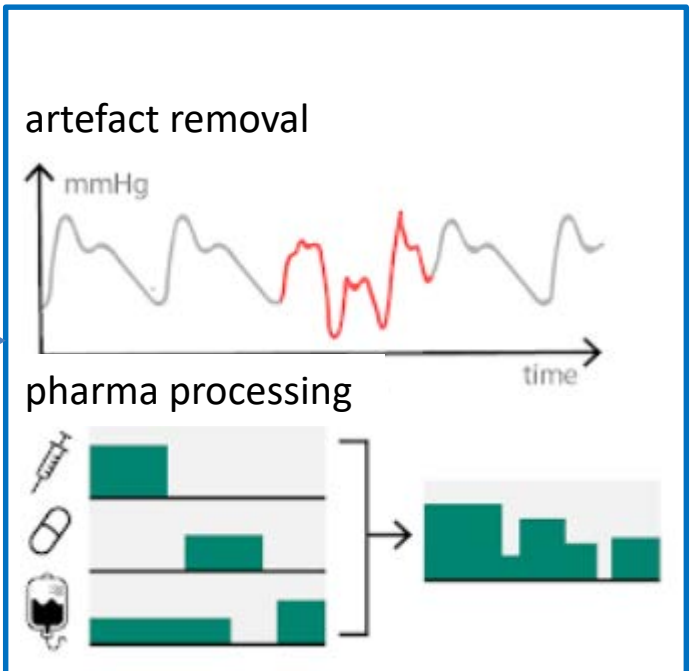


Data pre-processing

Preparation of raw data

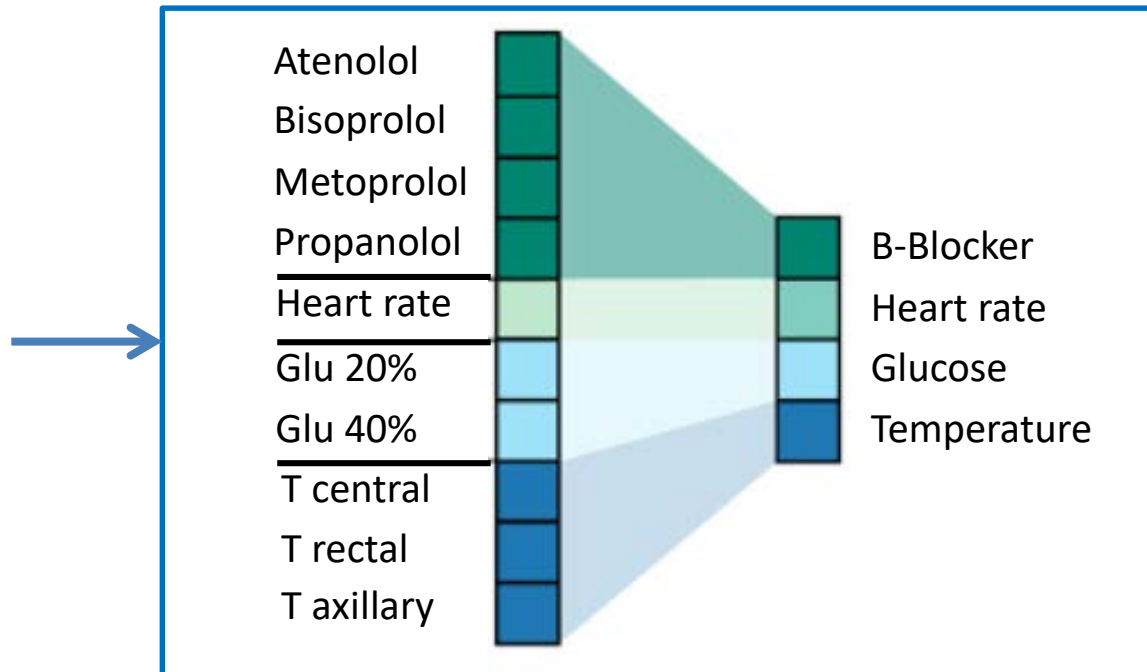


Data cleaning

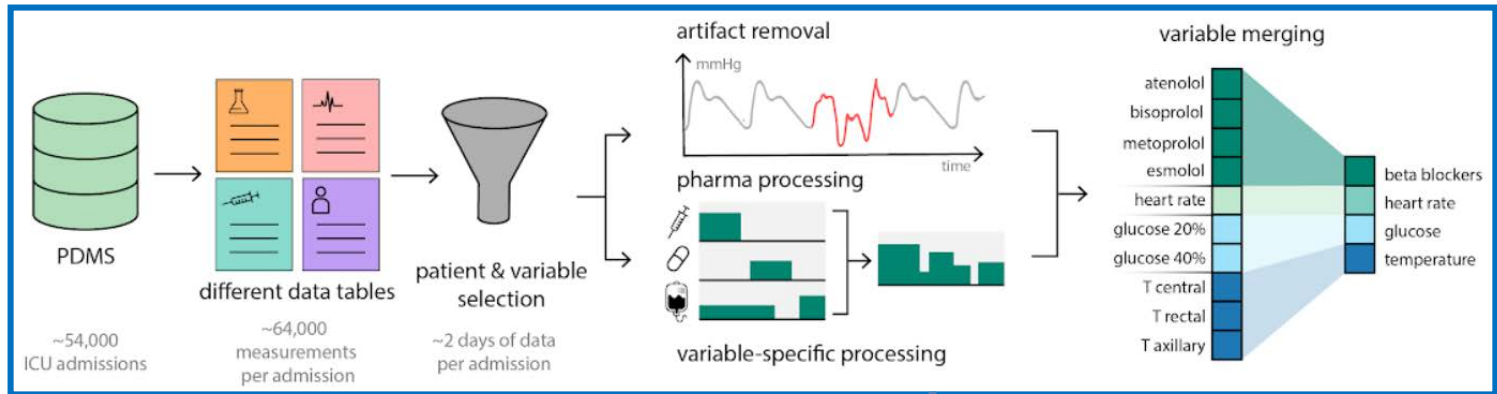


Data pre-processing

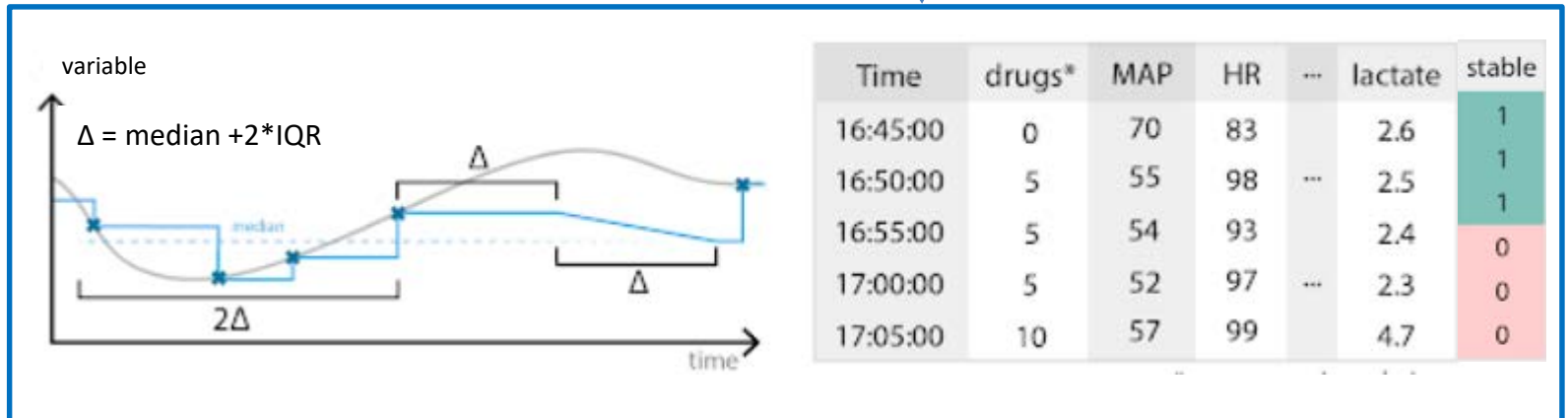
Variable merging



Data pre-processing



Adaptive imputation and label generation



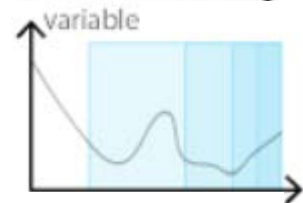
Feature generation



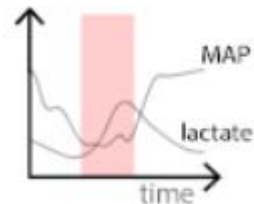
Static features: gender, age,



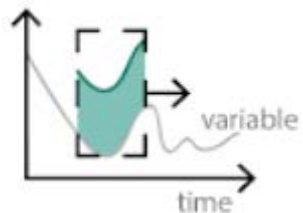
“Measurement-based” features: Time since the variable was last measured, and its value



Multi-resolution summary features: Summary statistics of the variable over different timescales

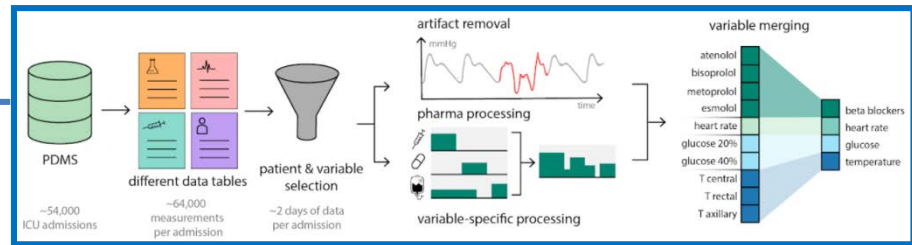


Patient stability status features: Check for partial fulfilment of criteria for an unstable patient state

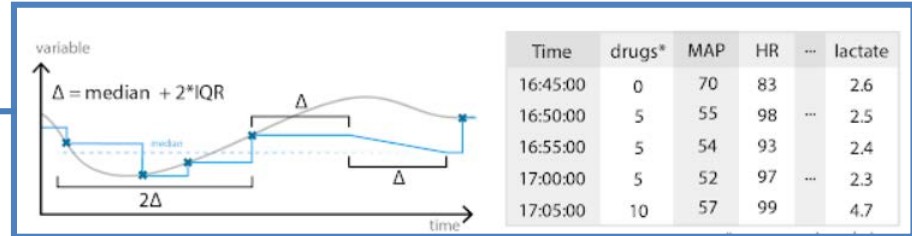


Shapelet features: represent distance to a given short time series (shapelet) which was shown to be indicative for onset of deterioration

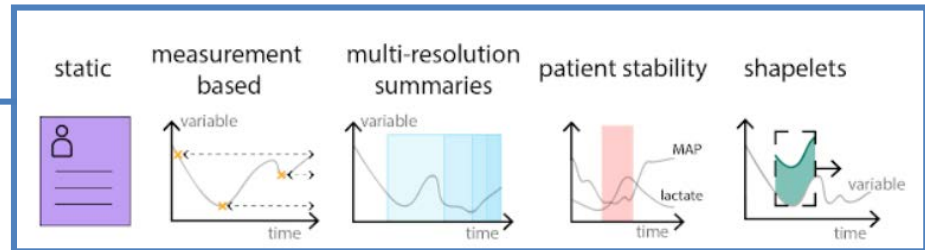
Preparation of raw data
Artefact removal
Variable merging



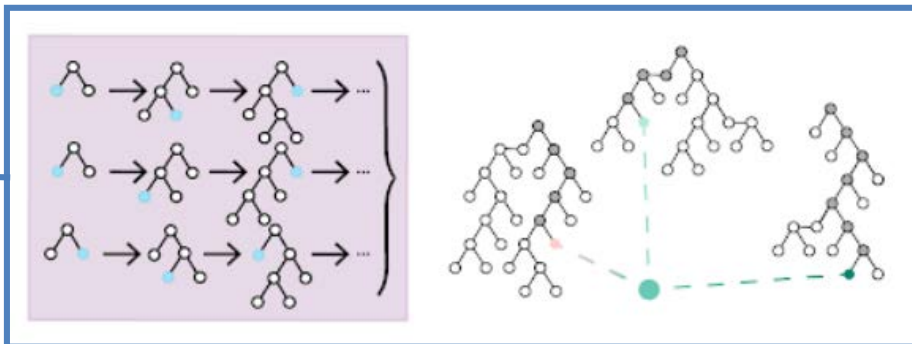
Imputation and label
(=endpoint) generation



Feature extraction



Model training



Barriers to use of PDMS data for research

- Legal and financial
 - Who owns the data and who are the stakeholders?
 - Who pays the project?
 - Are there other financial interests ?
- Ethical and legal aspects of use of patient data for research
 - Patient consent and IRB approval necessary?
 - Are patients identifiable? Can data be anonymized? Is the data coded?
- Technical
 - It's more complicated than you first think...