Temporal features for classification of patient-management patterns

Milos Hauskrecht

Michal Valko, Iyad Batal

Computer Science Department and Department of Biomedical Informatics, University of Pittsburgh

Outline

- High dimensional temporal datasets
- Building predictive models
- Temporal features

Models of patient-management decisions

Patient medical records today:

- Defined by complex multivariate time-series that consist of thousands of electronic entries related to patient conditions and treatments
- Potential to be utilized in predictions, decision support, discovery or acquisition of new clinical knowledge

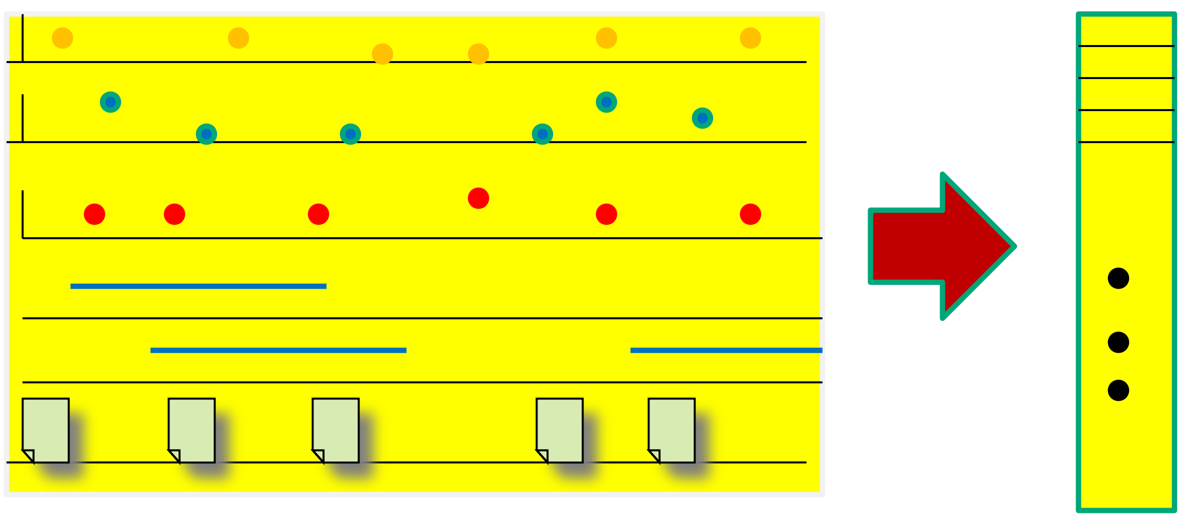
Objective of the work:

- Building of predictive models for patient-management decisions (lab and medication orders)
- **Key question:** What temporal features are the most important for predicting these orders?

Prediction models

Patient case is defined by complex multivariate time series, reflecting the history of test results, medications taken, demographics, vital signs, …

Patient case



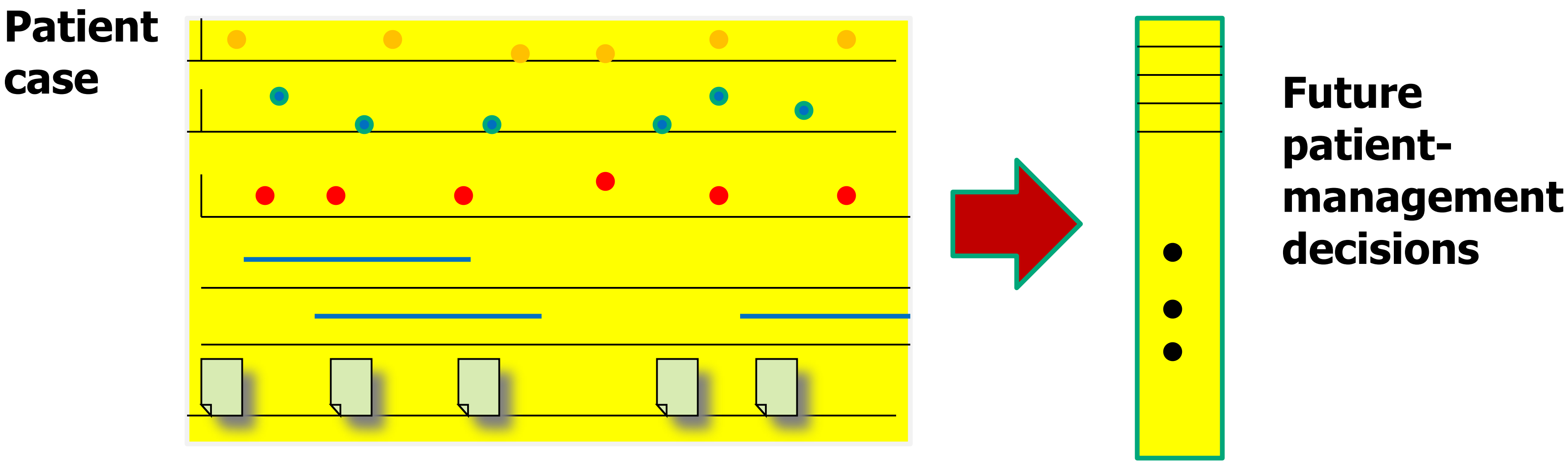
Future outcomes;

Future patient-management decisions

Predictions of: (1) future outcomes or (2) future patient-management decisions

Prediction models

Patient case is a complex multivariate time series, reflecting the history of test results, medications taken, demographics, vital signs, …



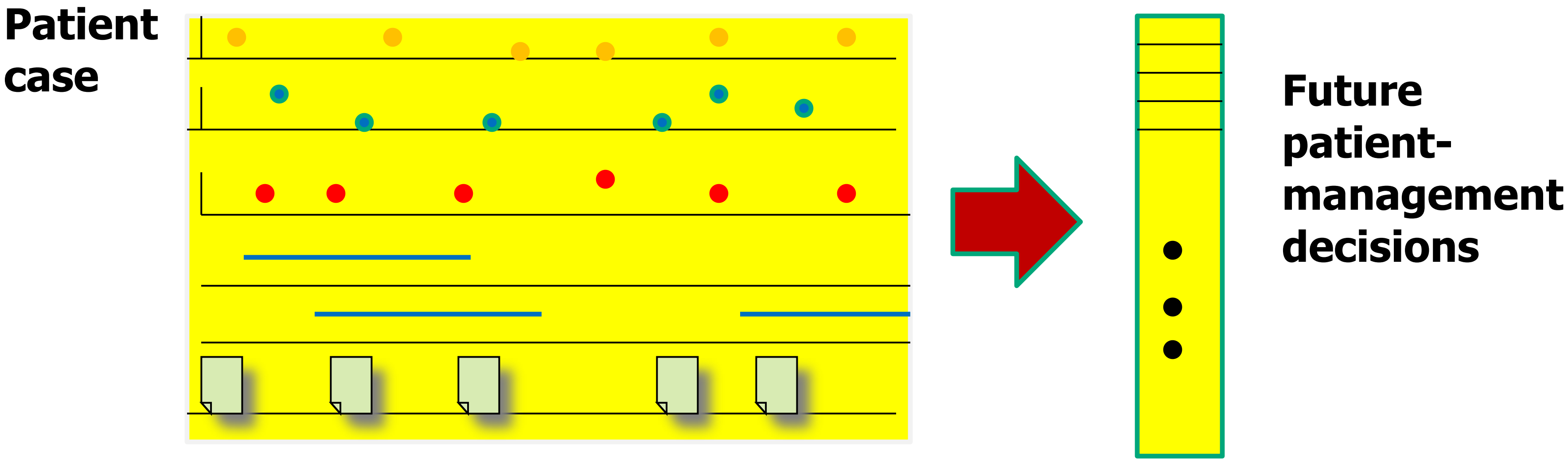
Predictions of future patient management decisions

- lab orders (e.g. order of liver function tests)
- medication orders (order warfarin, stop amiodarone)

Key challenges

- What temporal characteristics of the time series are most important for predicting patient-management decisions?

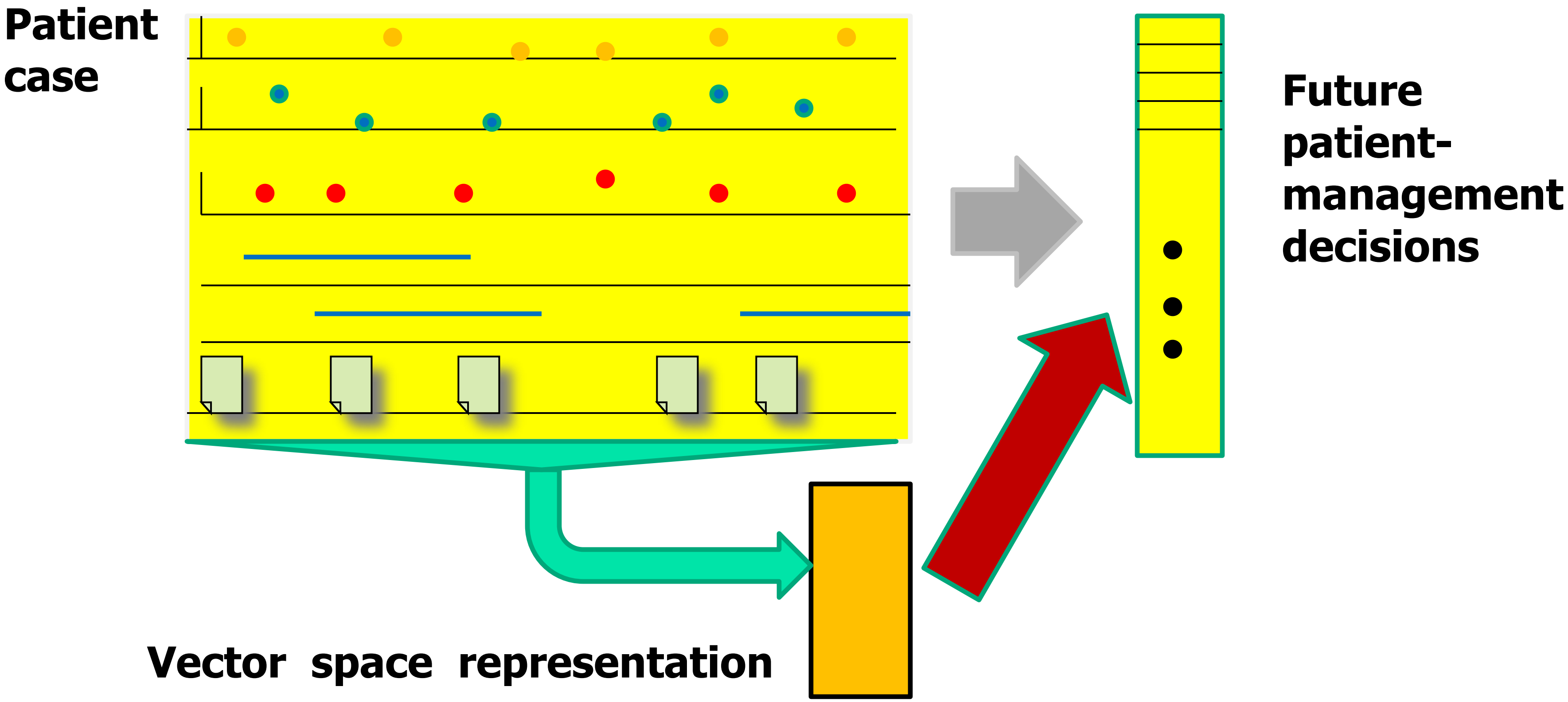
… or what are the temporal features to use in the model?



Key challenges

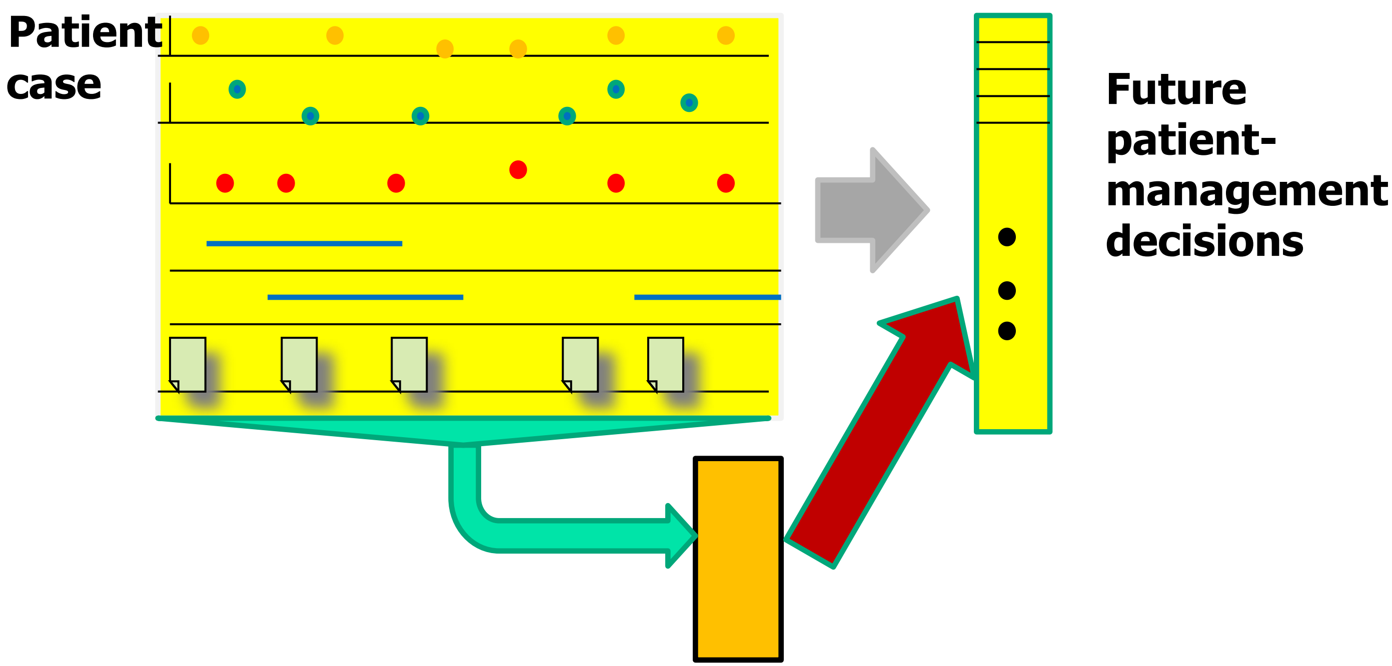
- What temporal characteristics of the time series are most important for predicting patient-management decisions?

… or what are the temporal features to use in the model?



Temporal features

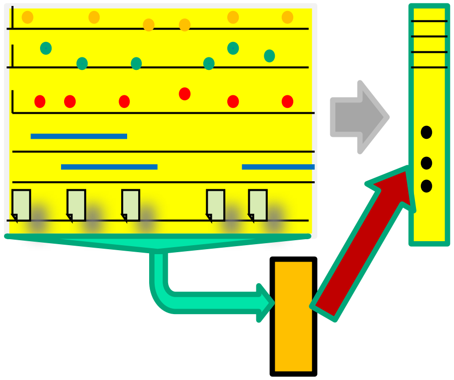
- A large number of temporal features candidates
- A small number of features to build a prediction model:
  - Decisions influenced by only few clinical variables
  - Necessary to prevent the over-fitting problem



Importance of temporal features

Objective:

- Understand what temporal features are important for predicting patient-management decisions



Methodology:

- Generate a large number of temporal features characterizing the patient state
- Assess univariate importance of temporal features for predicting different lab and medication orders

Data:

- Medication and lab order decisions for post-surgical cardiac patients

PCP database

Post-surgical cardiac patient (PCP) database

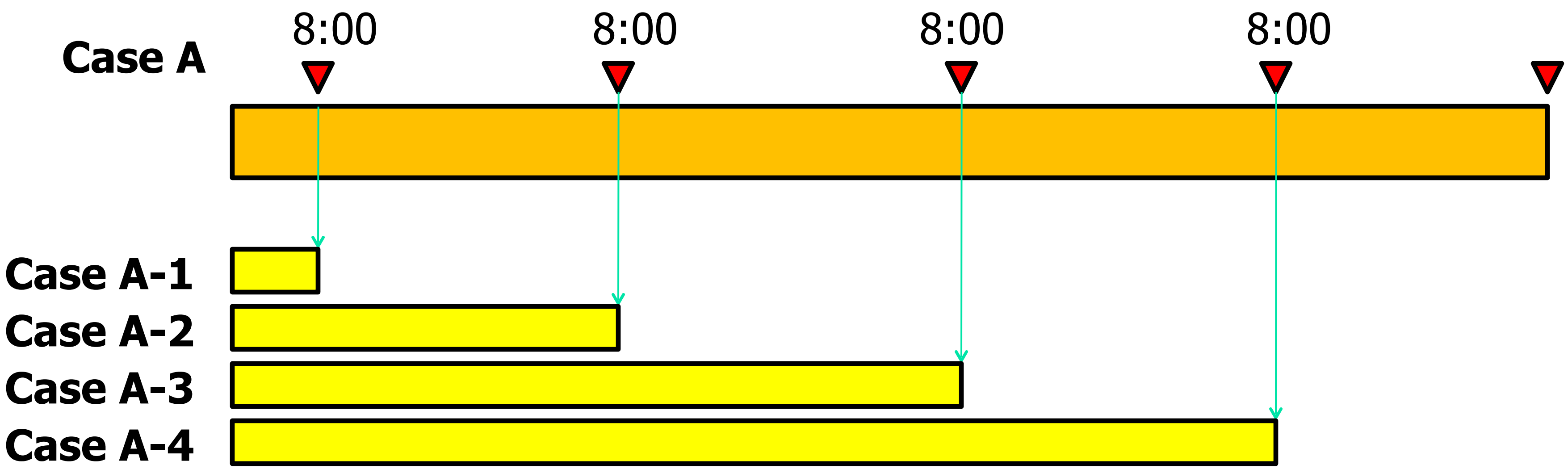
- A database of de–identified records for 4486 post–surgical cardiac patients treated at one of the University of Pittsburgh Medical Center (UPMC) teaching hospitals.

PCP database includes:

- All labs and tests (including standard and all special tests),
- two medication databases,
- Discharge summaries
- History &physical
- Progress reports
- EKG, radiology and special procedures, ER, OR reports,
- financial charges database.

Dataset used in the analysis

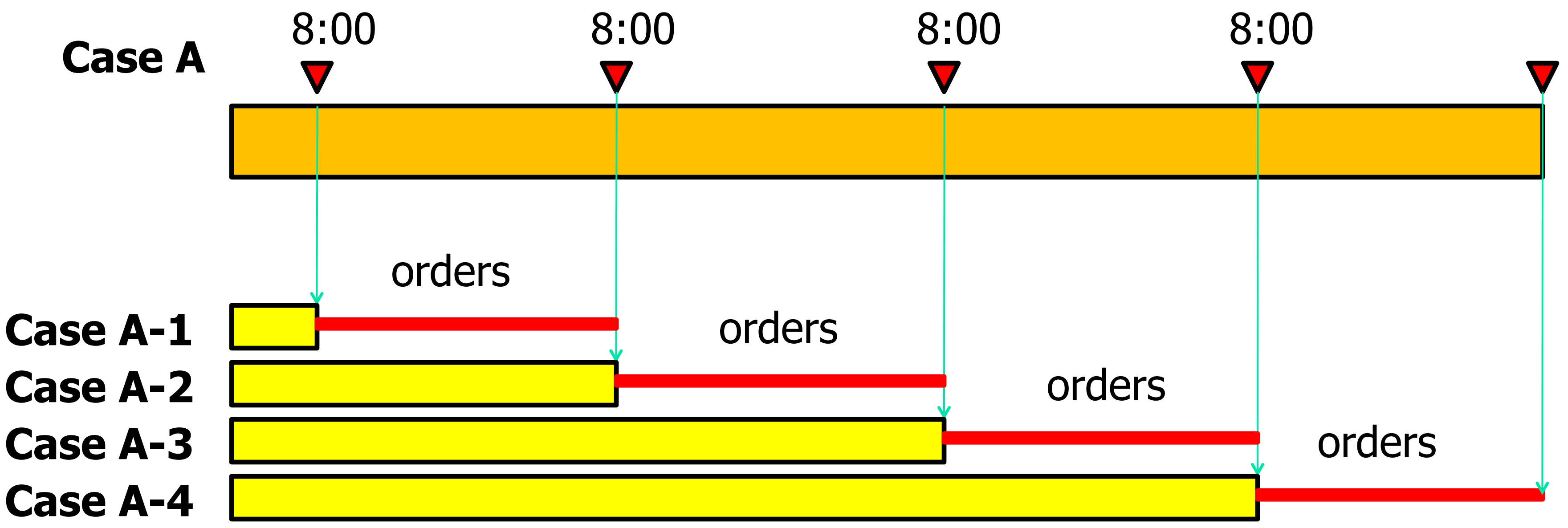
Patient records in PCP were segmented to generate multiple patient case examples



- Fixed time (8:00am) segmentation of the electronic health record
- 4486 EHRs used to obtain 30,828 patient examples

Dataset

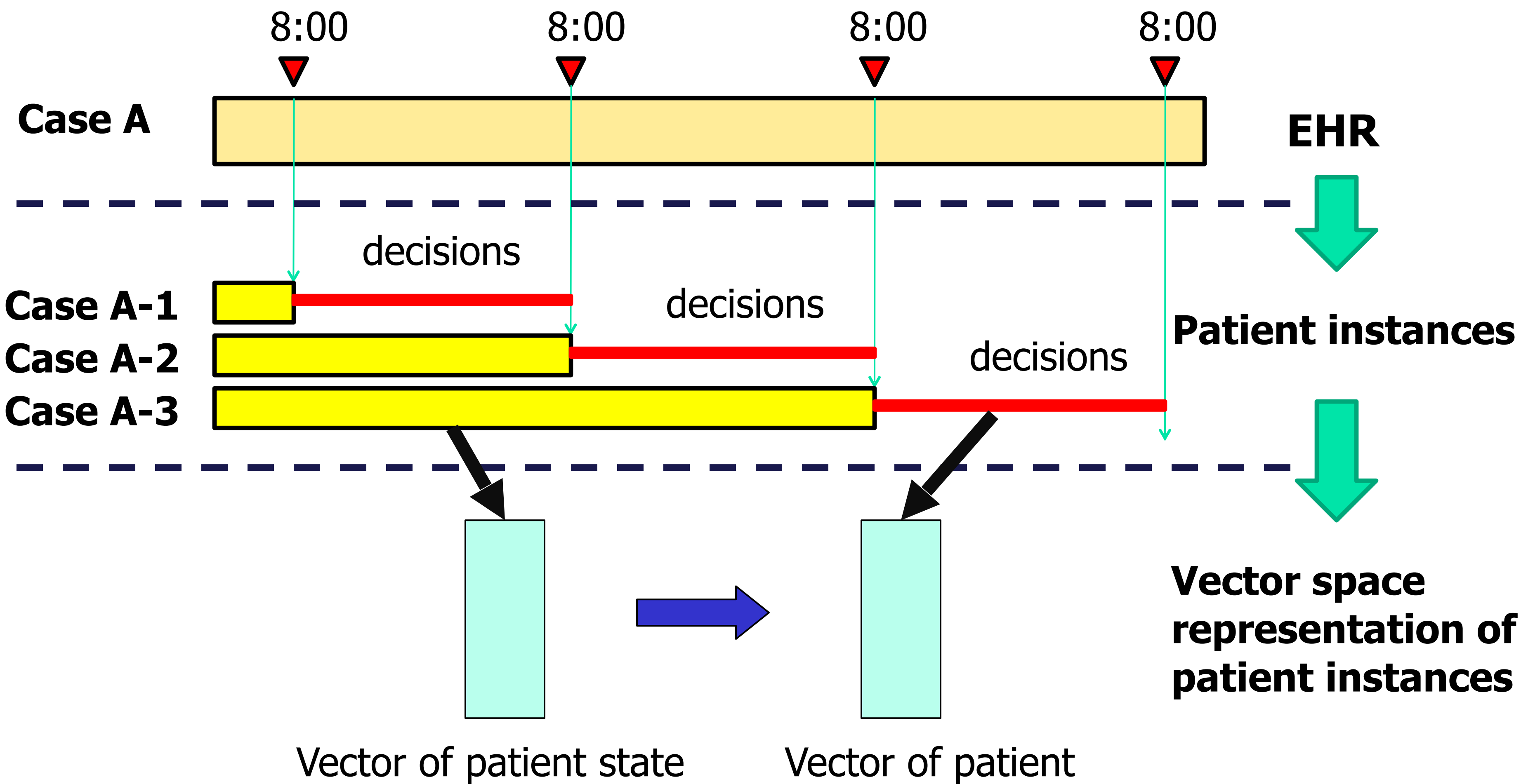
Every example after segmentation was associated with lab and medication orders done in next 24 hours



- 335 lab orders
- 407 medication orders

Dataset

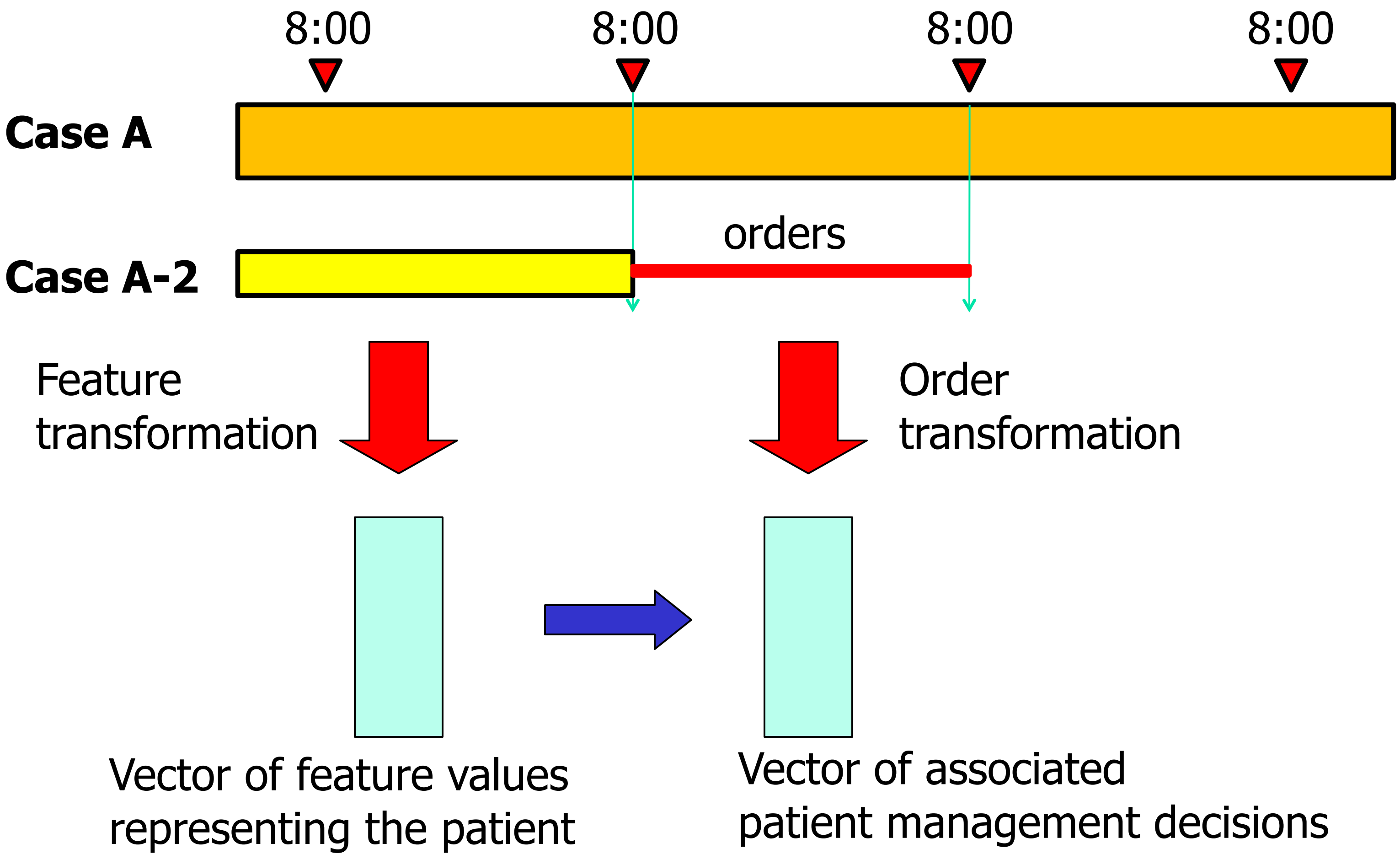
N



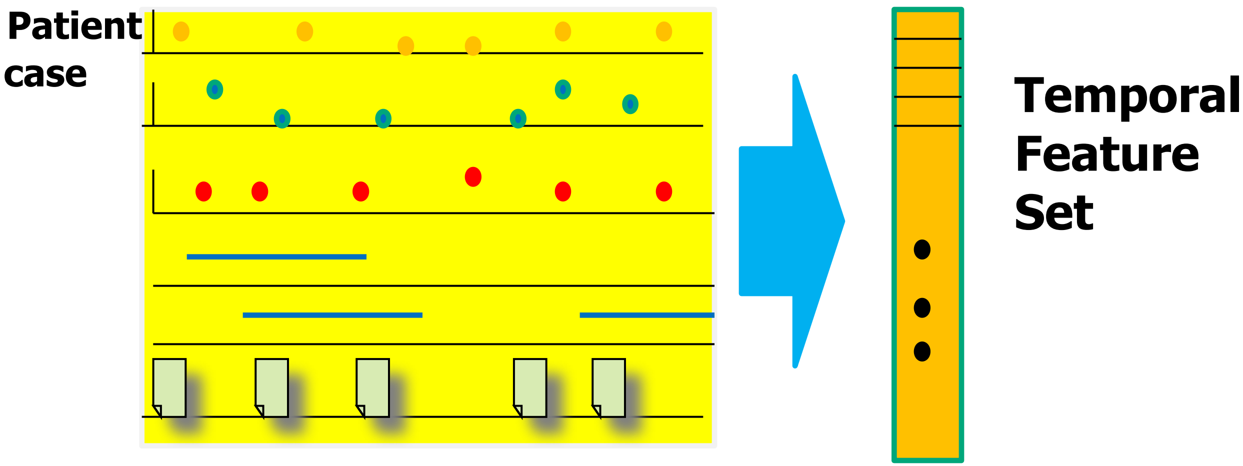
8:00
8:00
8:00
8:00
Case A
EHR
decisions
Case A-1
decisions
Case A-2
Patient instances
decisions
Case A-3
Vector space representation of patient instances
Vector of patient state
Vector of patient

Vector space representation

Data for every patient state instance and patient management decisions associated with it are then transformed into a vector space representation.



Temporal features



- **Five feature groups:**
  - Lab features (335 labs, 40/4 features per lab)
  - Medications (407 medications, 4 features per med)
  - Demographics (total of 3 features)
  - Procedures (31 procedures)
  - Heart support devices (5 device types, 2 features per device)

Temporal features for labs

Lab features:

- Discrete labs (4 features per lab)
- Numerical labs (40 features per lab)

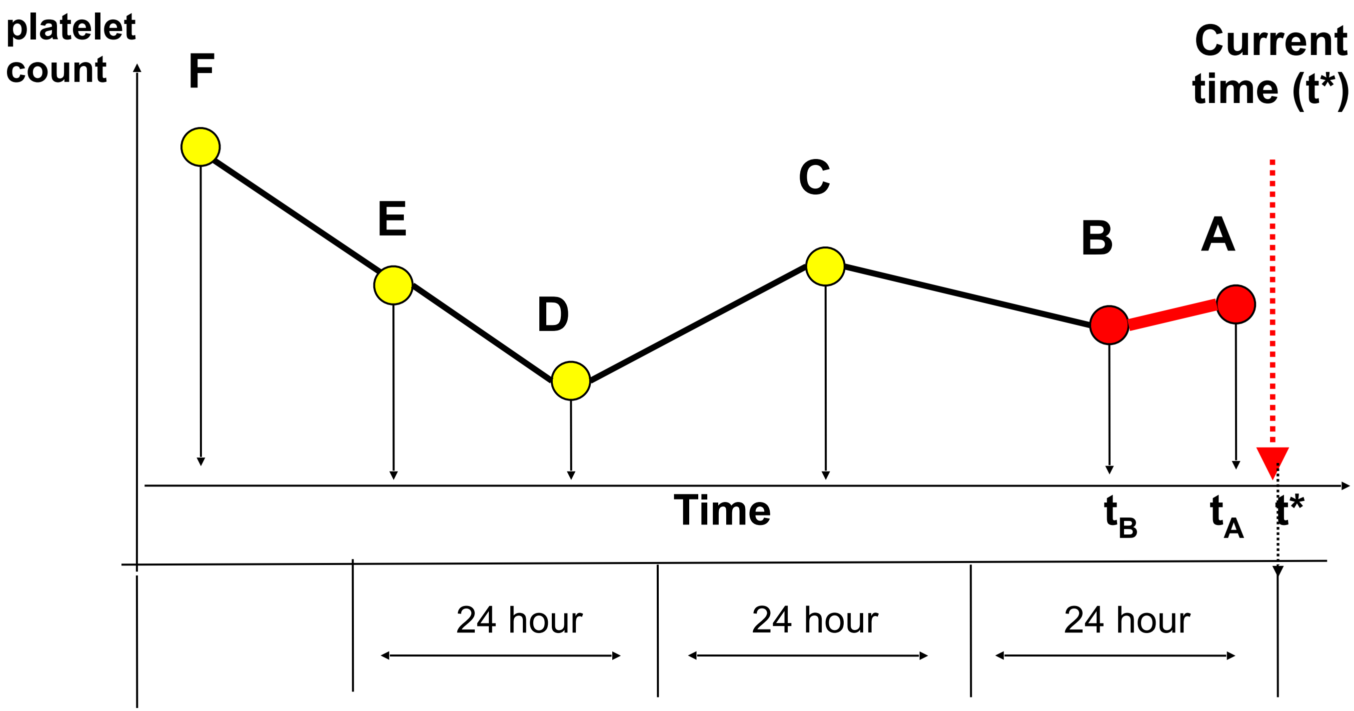
Common features (for categorical and numerical labs):

- Lab value known
- Time since last value
- Last value (numeric value or one of the categorical values)

Additional feature for categorical labs:

- Indicator of a change from the normal (default) value (e.g. NEG → POS)

Features for numerical labs



Temporal features:

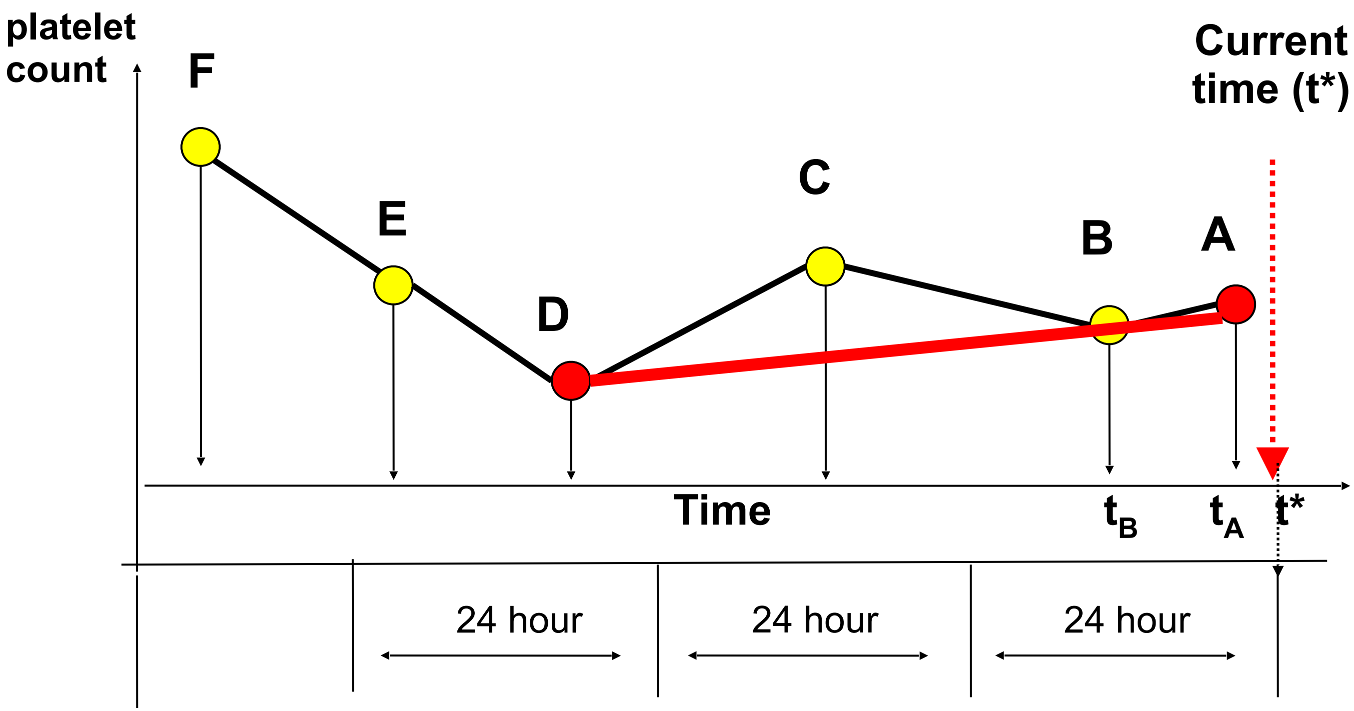
Last value: A

Last value difference = B-A

Last % change = (B-A)/B

Last slope = (B-A) / (t_B-t_A)

Features for numerical labs



Temporal features:

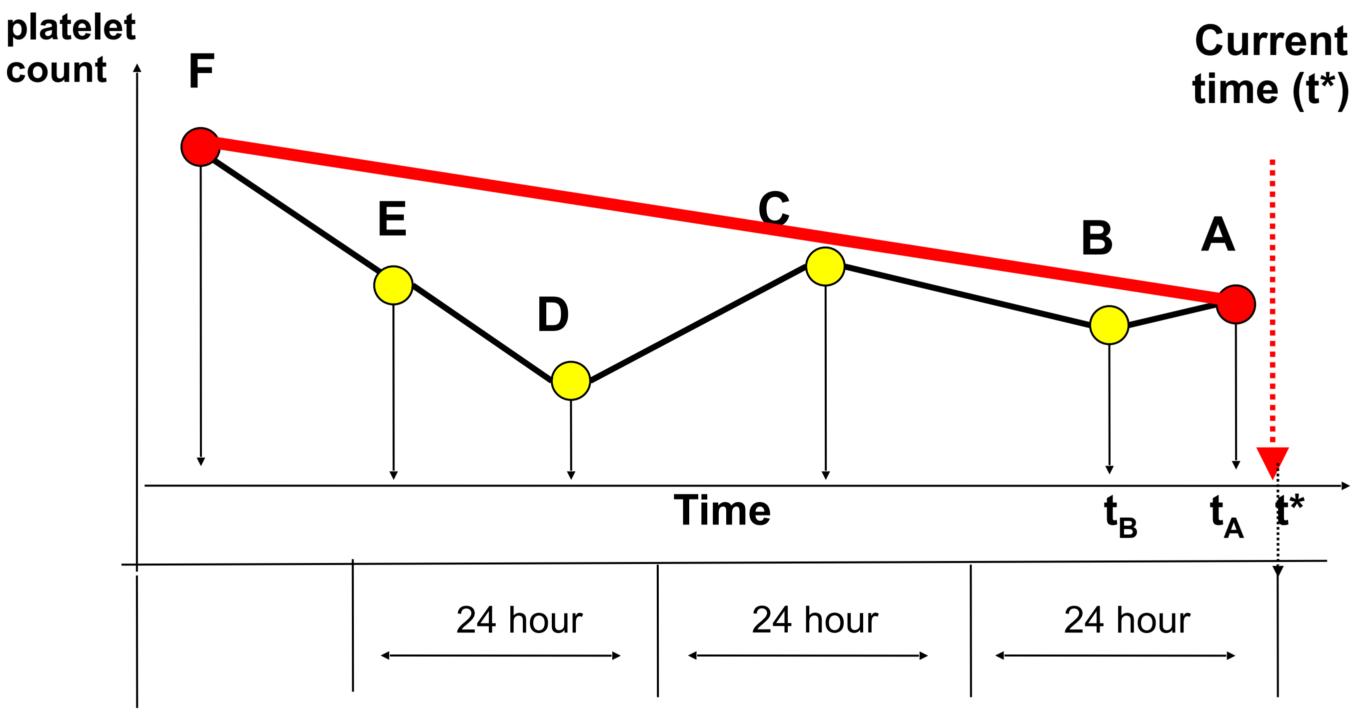
Nadir = D

Nadir difference = A-D

Nadir % difference = (A-D)/D

Nadir slope= $(A\text{-}D)/(t_D\text{-}t_A)$

Features for numerical labs



Temporal features:

Baseline = F

Baseline difference = F-A

Baseline % difference = (F-A)/F

Baseline slope= (A-D)/ (t_F-t_A)

Features for numerical labs

Other numerical lab features:

- absolute difference of the last value from the normal range midpoint
- Horizon features – similar to nadir

  Horizon value, Horizon difference, Horizon % difference, Horizon slope= $(A-D)/(t_D-t_A)$
- 24 hour lab value averages

Temporal features

Medication features:

- 407 medications
- Dosages were not considered
- 4 features per medication:
  - Indicator if it is the current medication
  - Time on current medication
  - Time since the first order of the medication
  - Time since last change

Demographics:

- Total 3 features

Temporal features

Procedure features:

- 31 procedures
- For every procedure we generate:
  - Procedure indicator:
  - Time since the procedure
  - Times since the first procedure (for repeats)

Heart-support device features:

- 5 device groups (pacemaker, ECMO, etc)
- For every group:
  - Indicator if the patient is supported by the device
  - Time on the device

Temporal features

Procedure features:

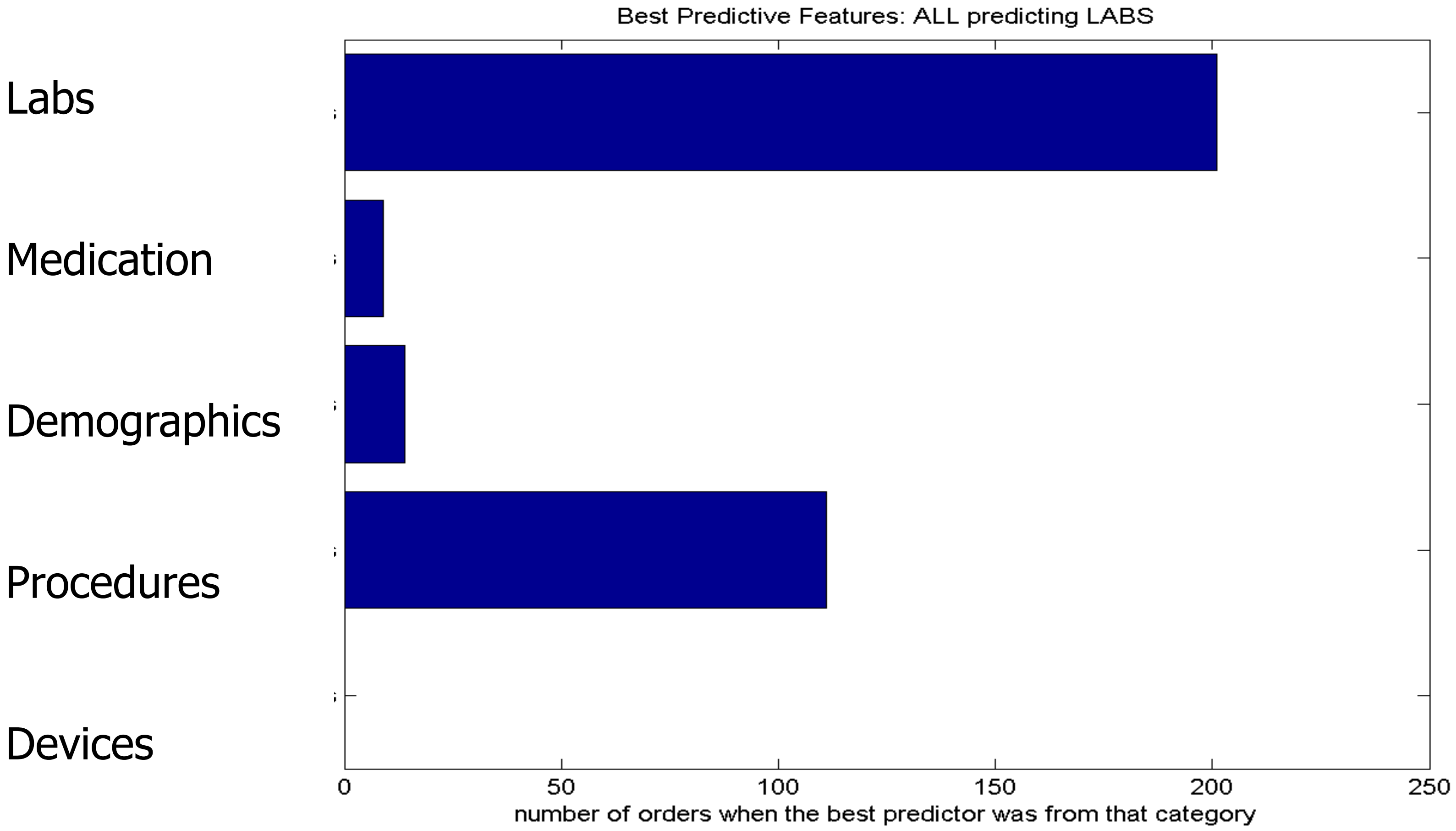
- 31 procedures
- For every procedure we generate:
  - Procedure indicator:
  - Time since the procedure
  - Times since the first procedure (for repeats)

Heart-support device features:

- 5 device groups
- For every group:
  - Indicator if the patient is supported by the device
  - Time on the device

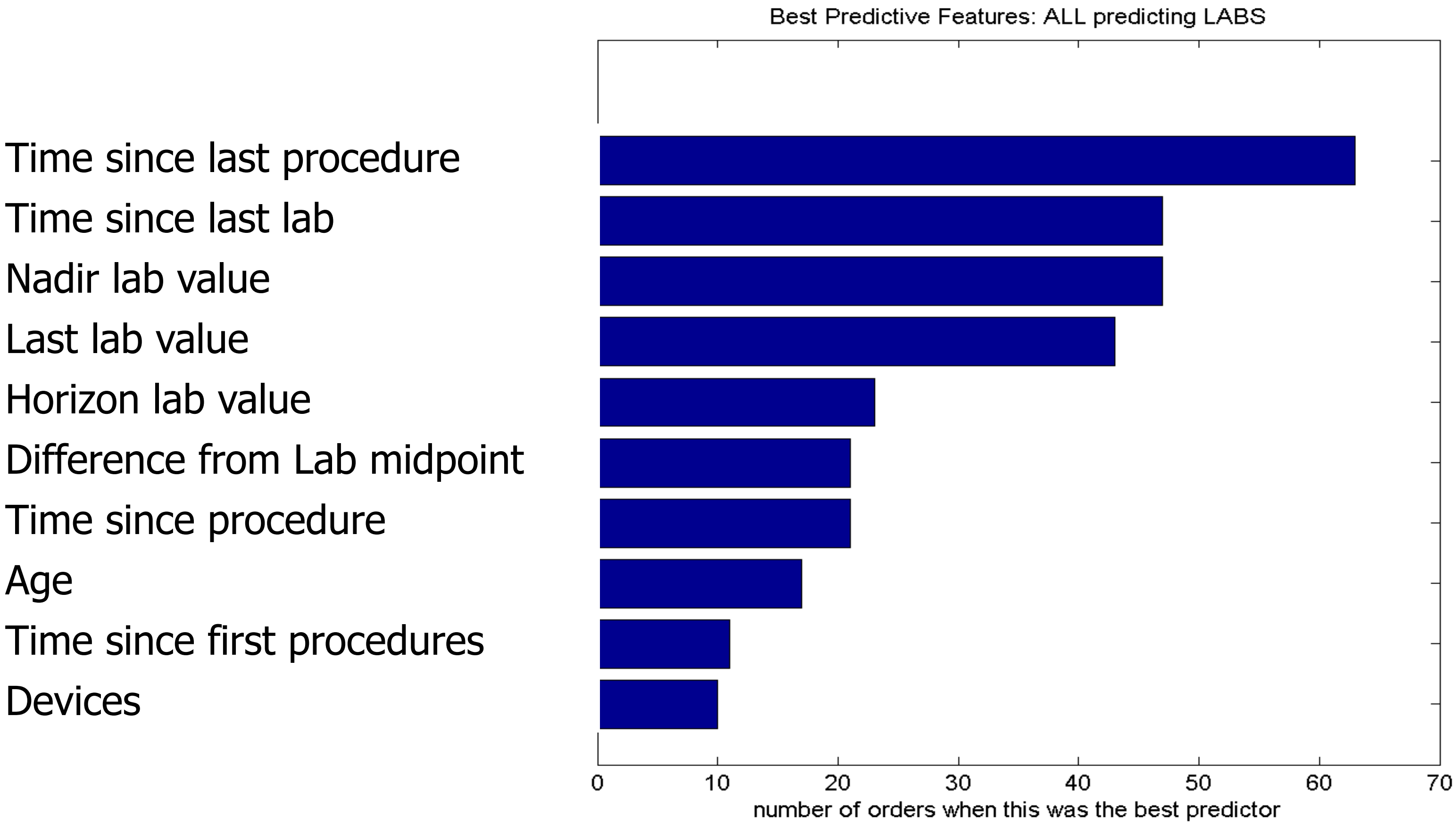
Importance of temporal features

5 feature groups predicting lab orders



Importance of temporal features

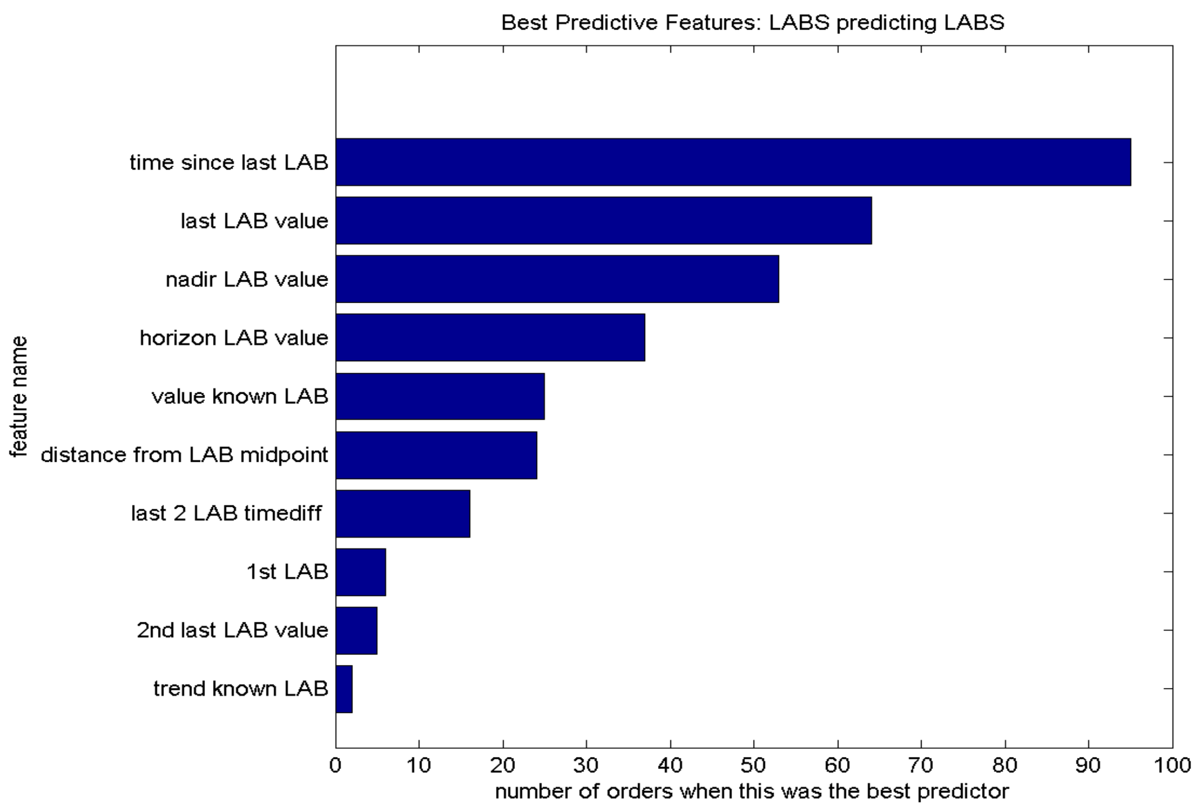
Lab orders versus all features



Dependencies

Lab orders versus the same lab features

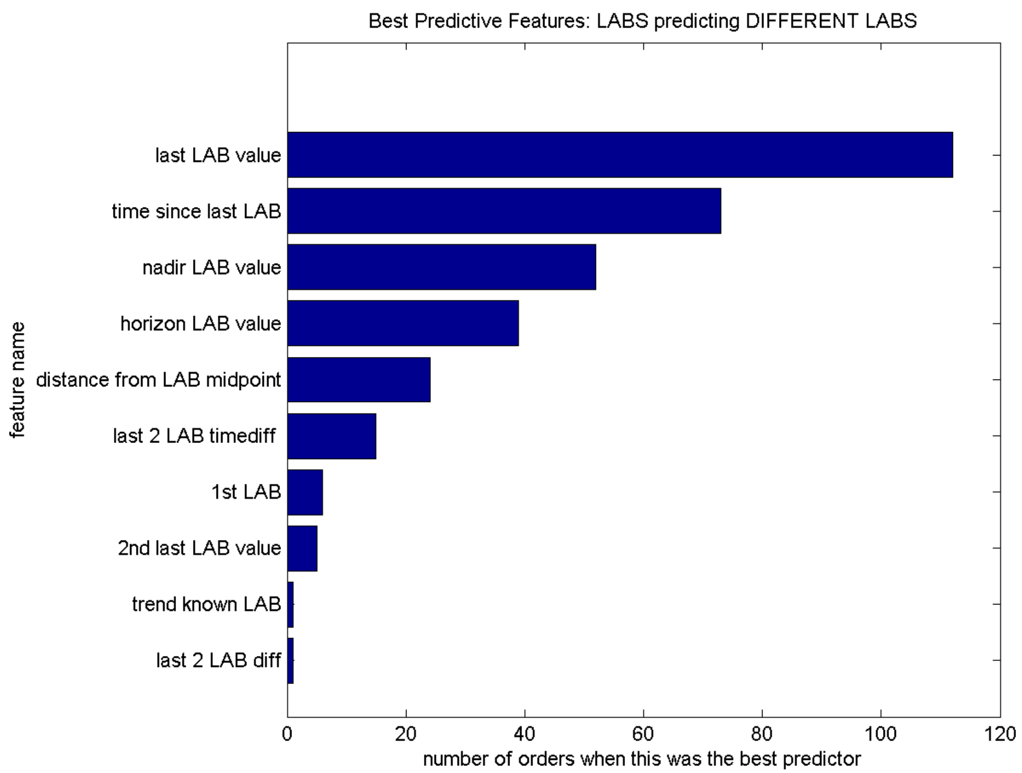
(e.g. GLU → GLU order, PLT → PLT orders)



Dependencies

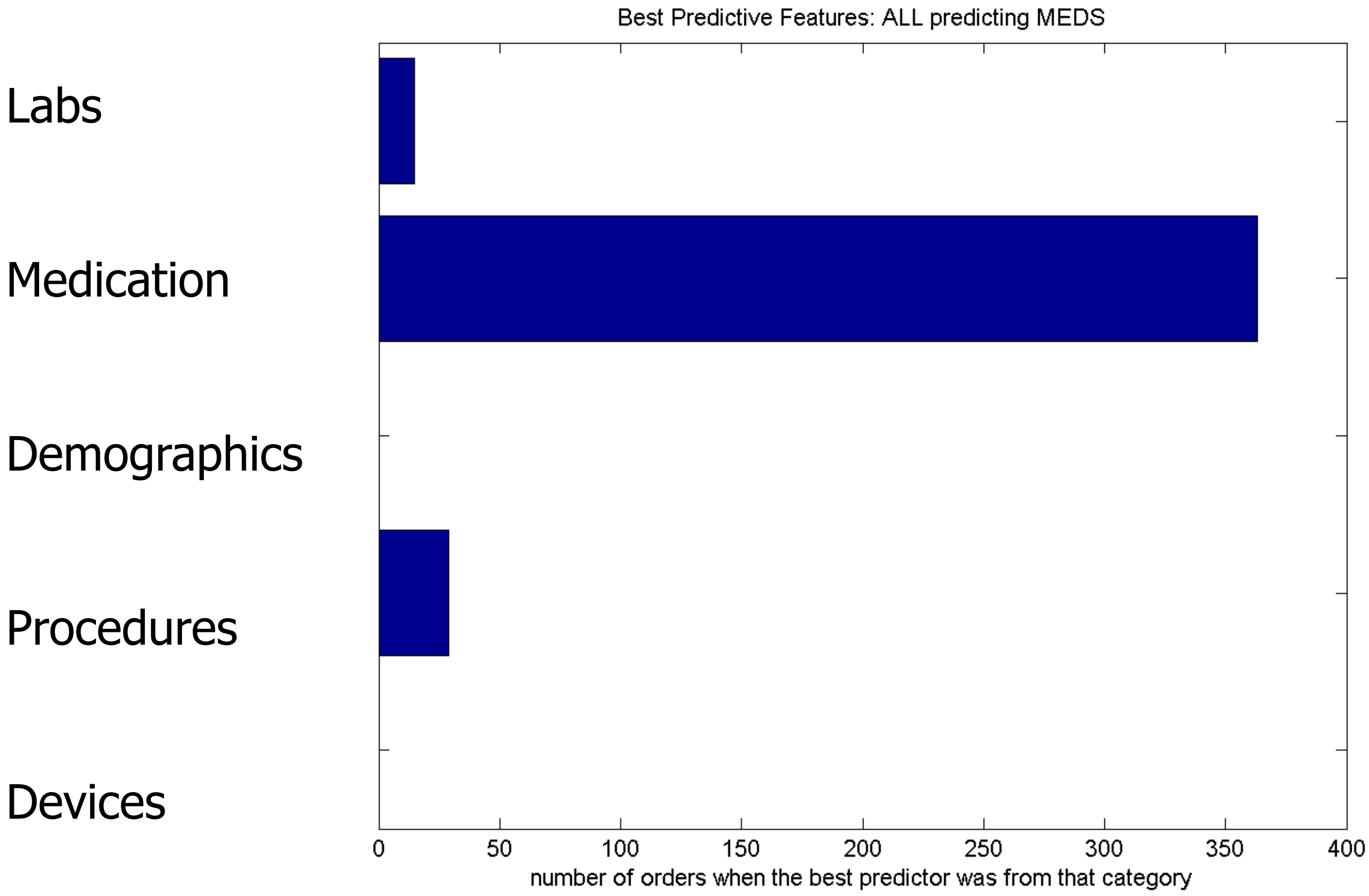
Lab orders versus the lab features for other labs

(e.g. GLU → GLU order, PLT → PLT orders)



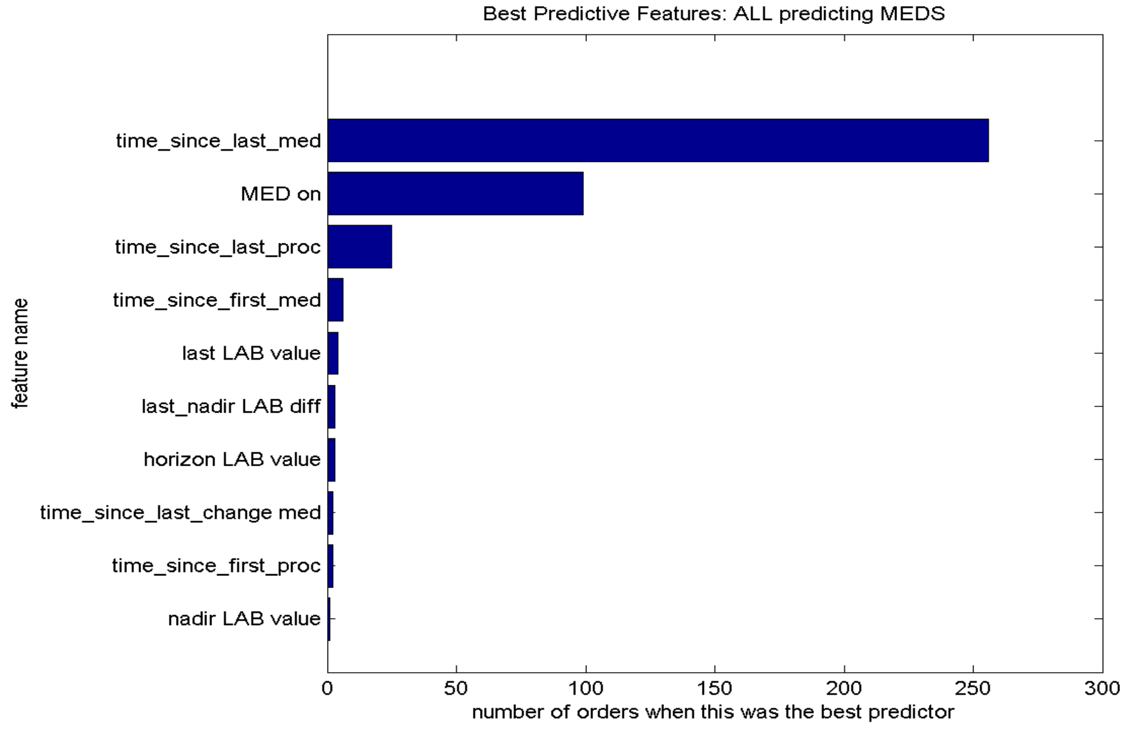
Importance of temporal features

Med orders versus the feature categories



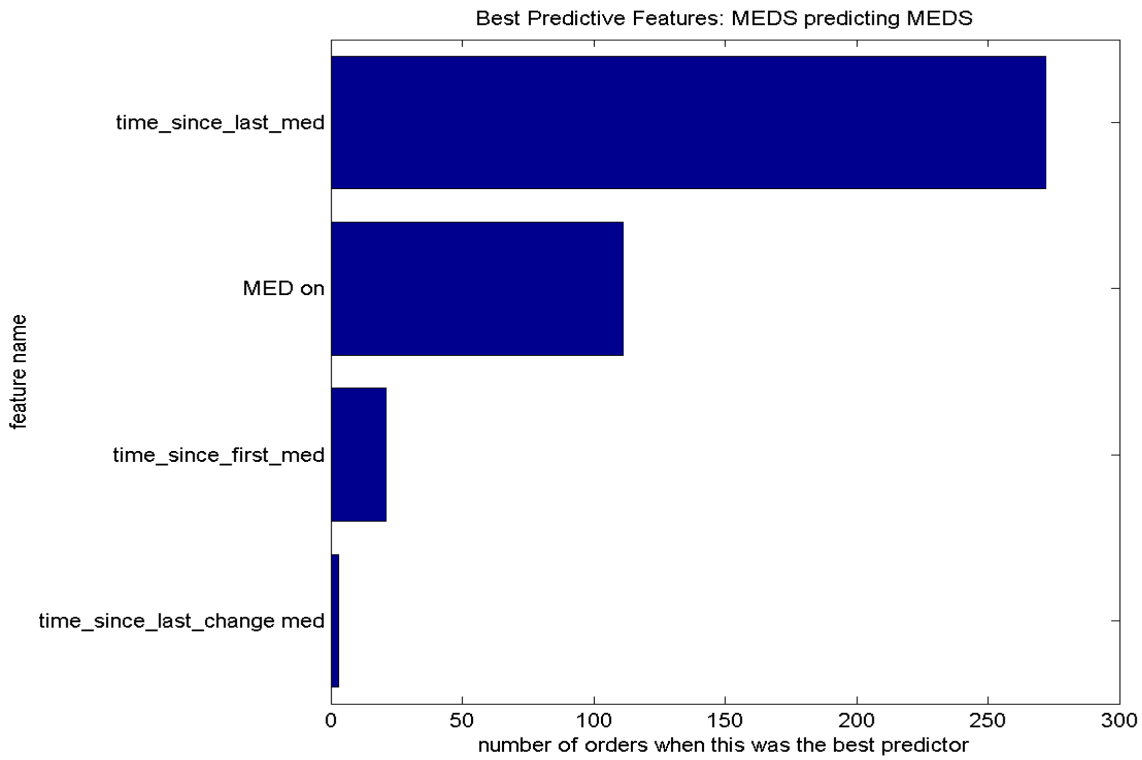
Dependencies

Med orders versus all features



Dependencies

Med orders versus med features



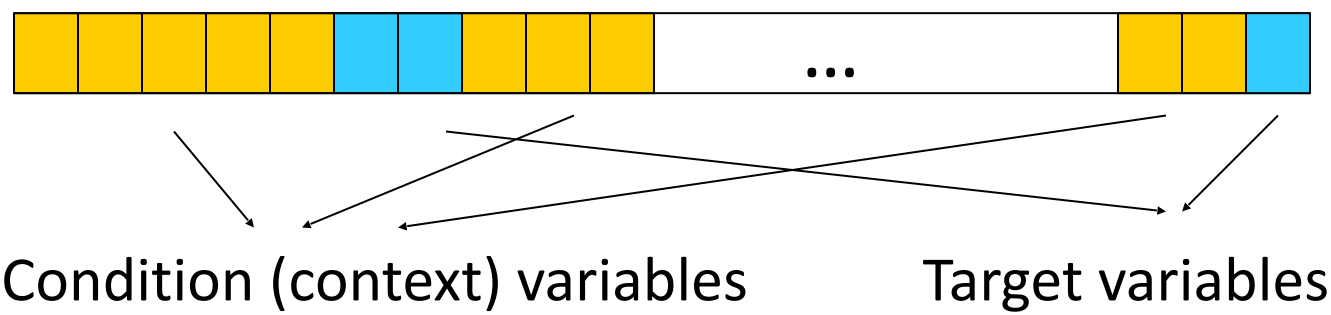
Dependencies

Med orders versus features for the same medication

Conditional anomaly detection

Conditional anomaly detection (CAD):

- Methods for detecting the presence of unusual data patterns depending on the values of other data features
- Data entry:



Question we ask:

- Given the values of context variables are the values of target variables unusual?

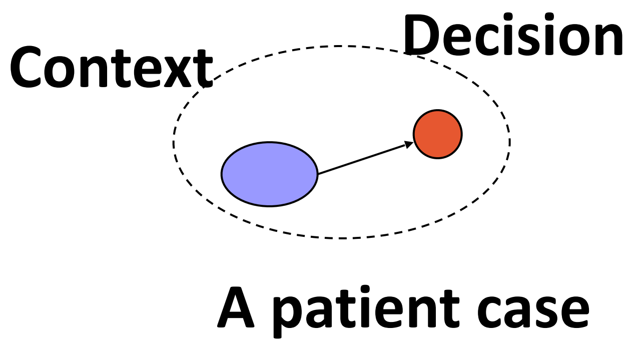
Conditional anomaly detection

Our goal:

- **Detect unusual patient management decision**

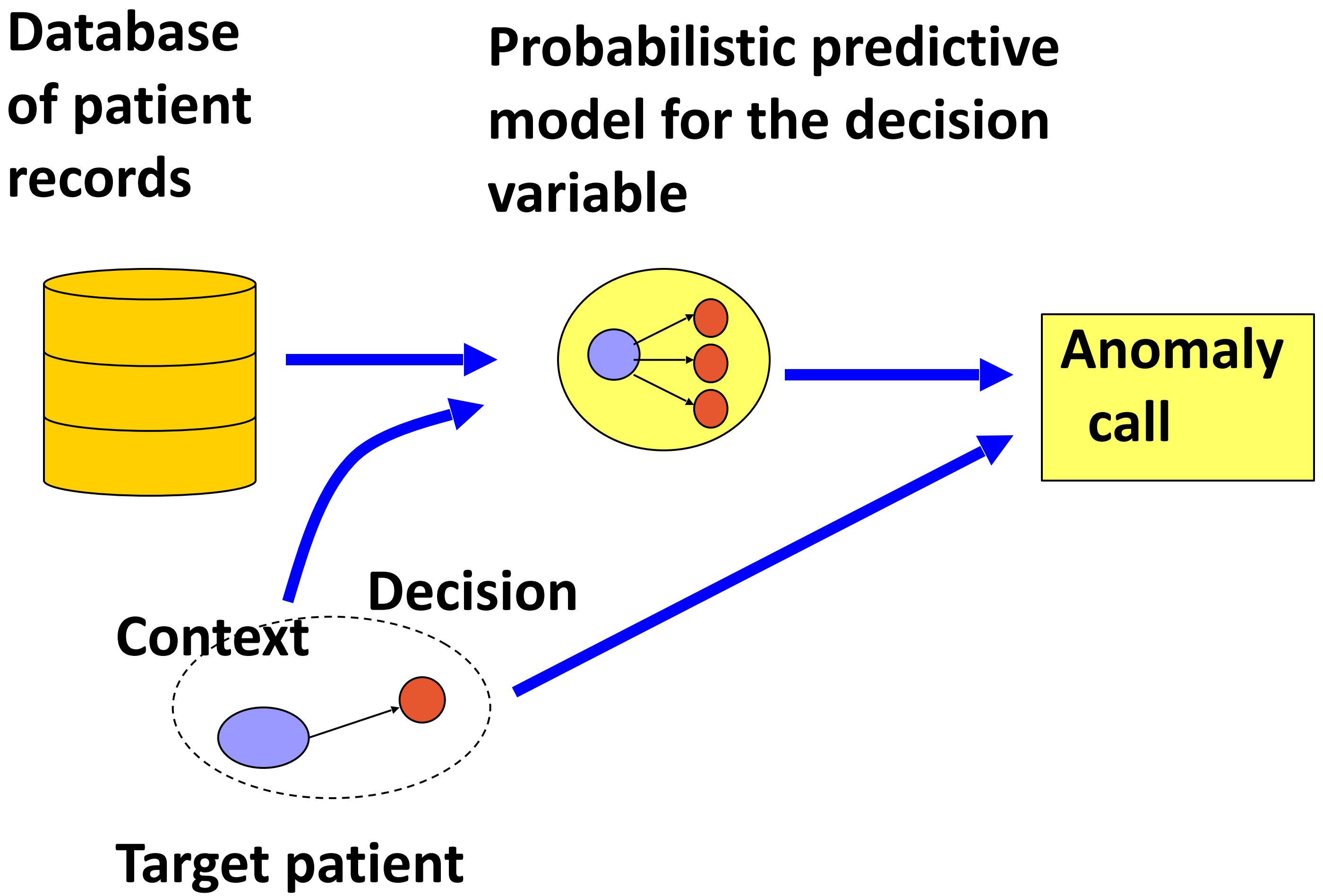
Formulation of the problem in CAD:

- **Condition (context) variables:** define the patient state
- **Target variable:** a patient-management decision



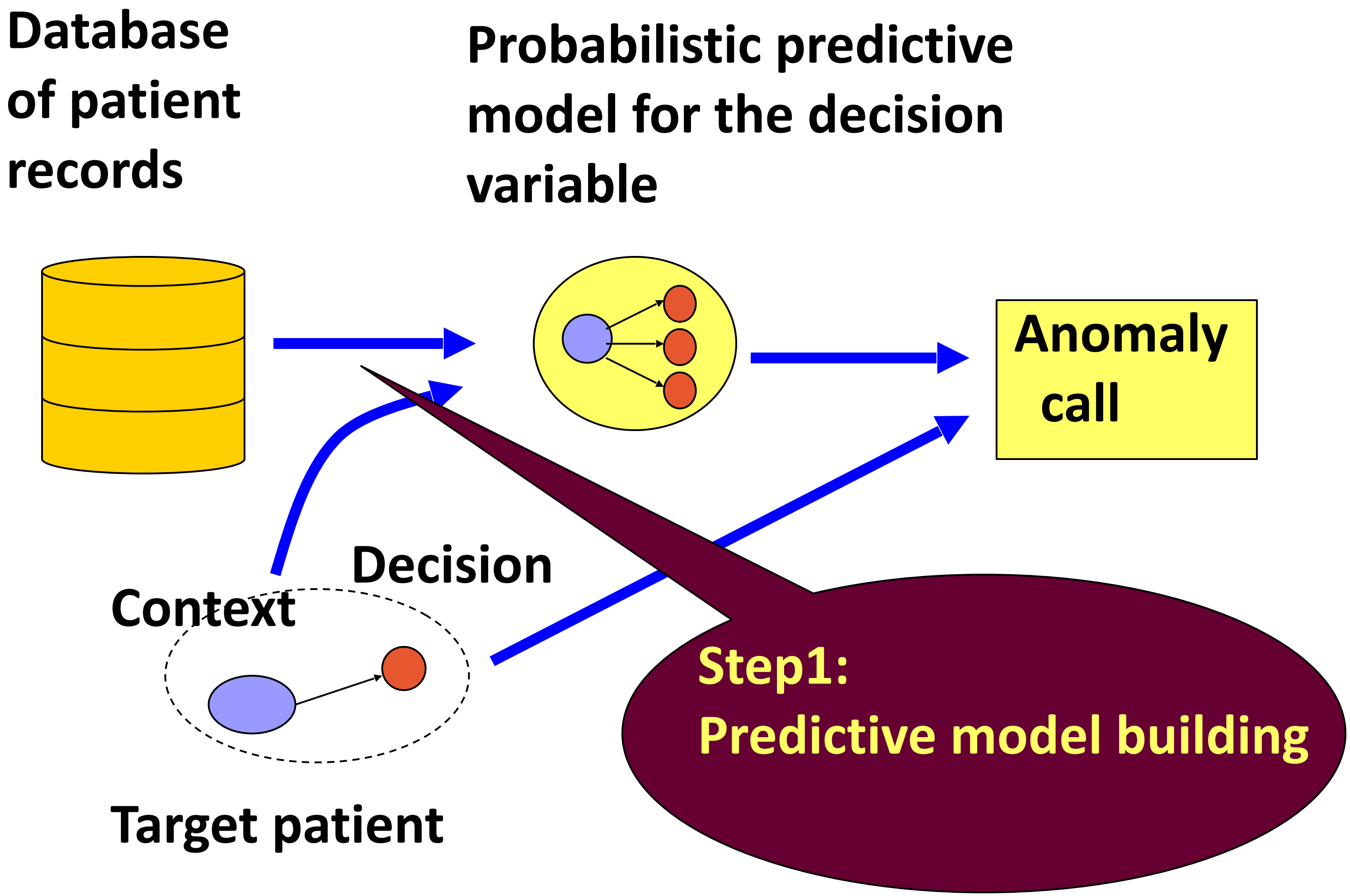
Conditional anomaly detection

- **Probabilistic conditional anomaly detection**



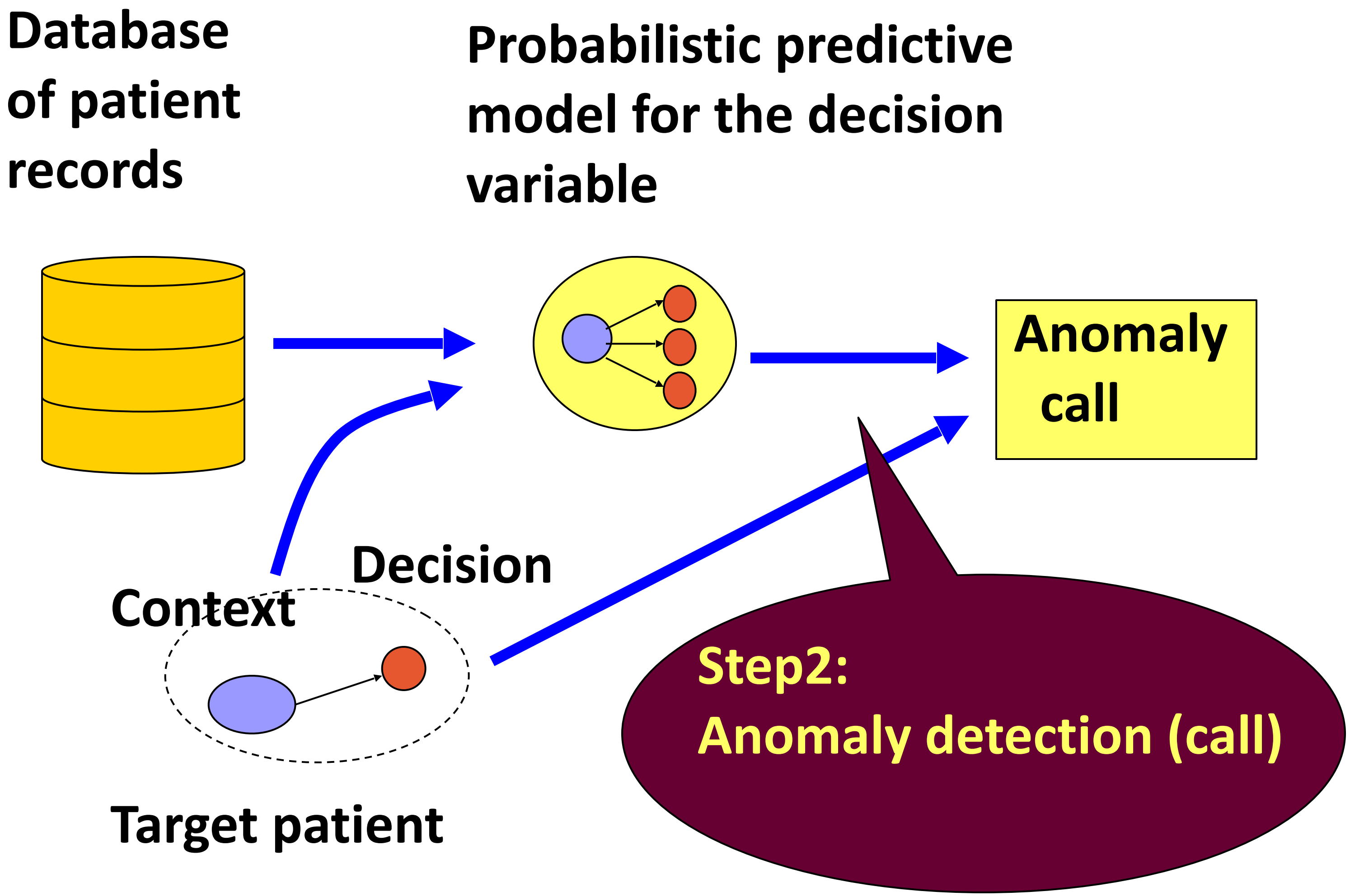
Conditional anomaly detection

- **Probabilistic conditional anomaly detection**



Conditional anomaly detection

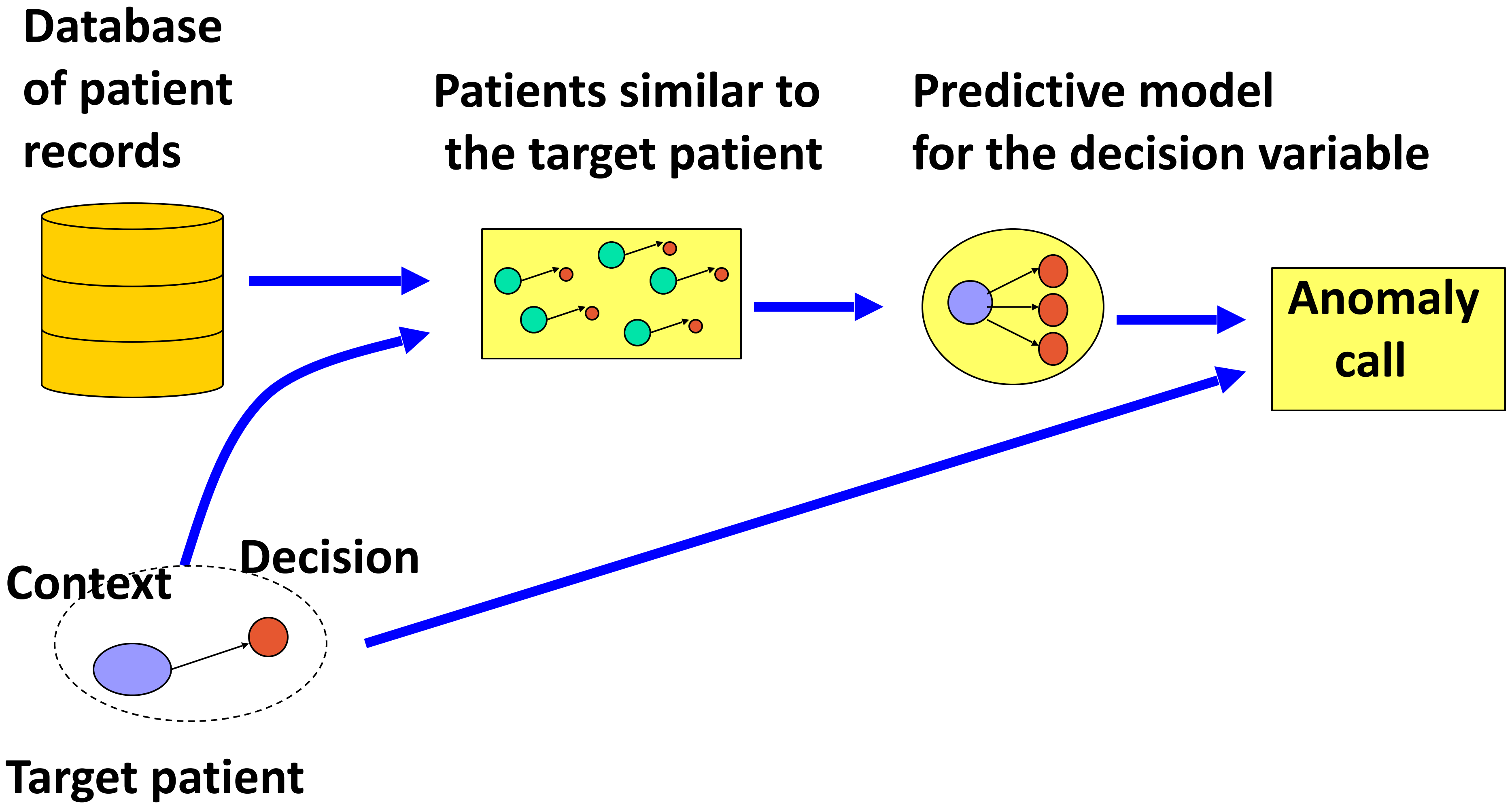
- **Probabilistic conditional anomaly detection**



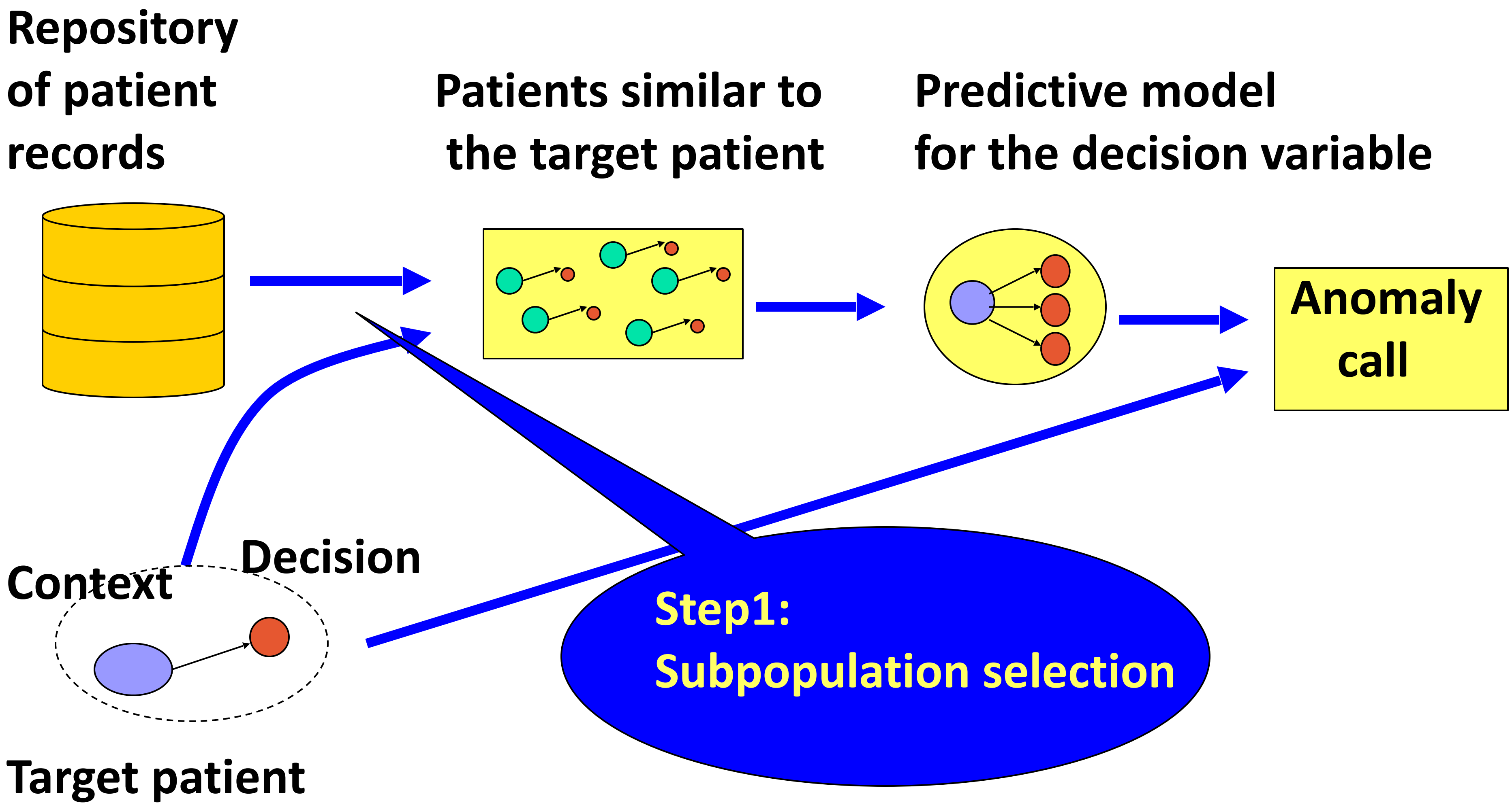
Probabilistic conditional anomaly detection

- **Predictive model:**
  - Learned from the database of past patients
  - Defines a distribution of decisions for the current patient with respect to the past patient
- A task very similar to the classification learning problem but we are trying to learn an accurate **predictive distribution**

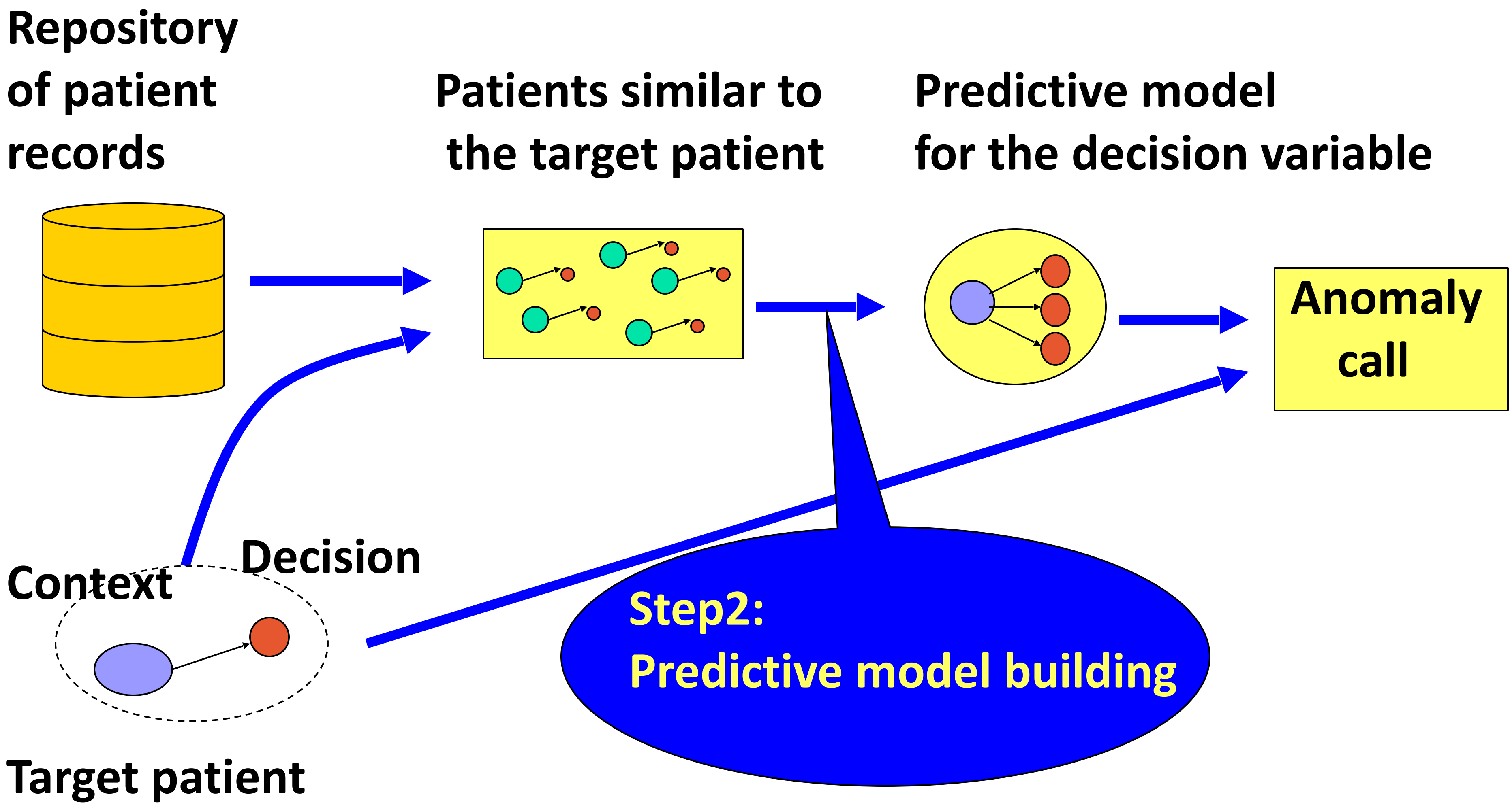
Conditional anomaly detection



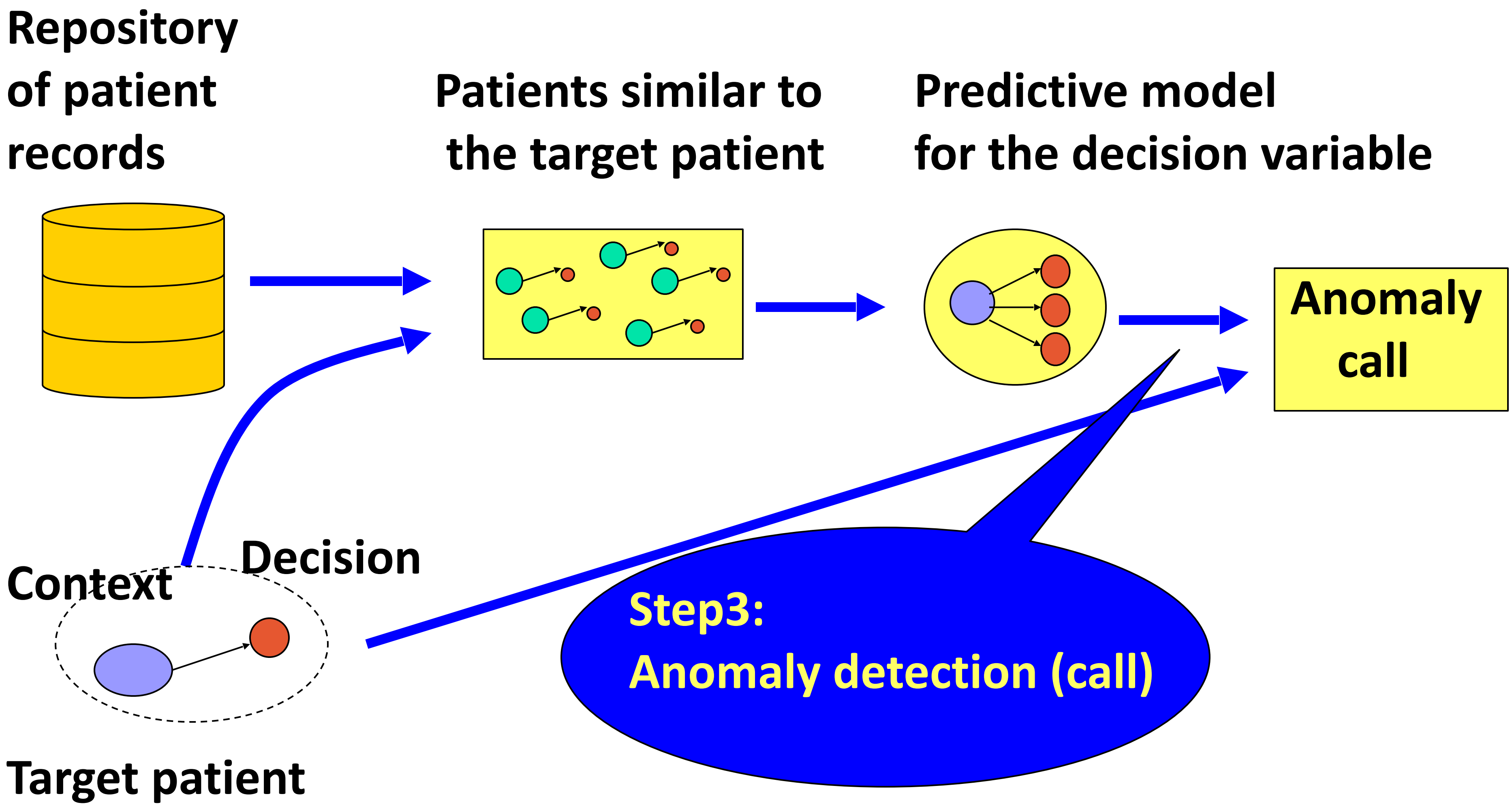
Conditional anomaly detection



Conditional anomaly detection



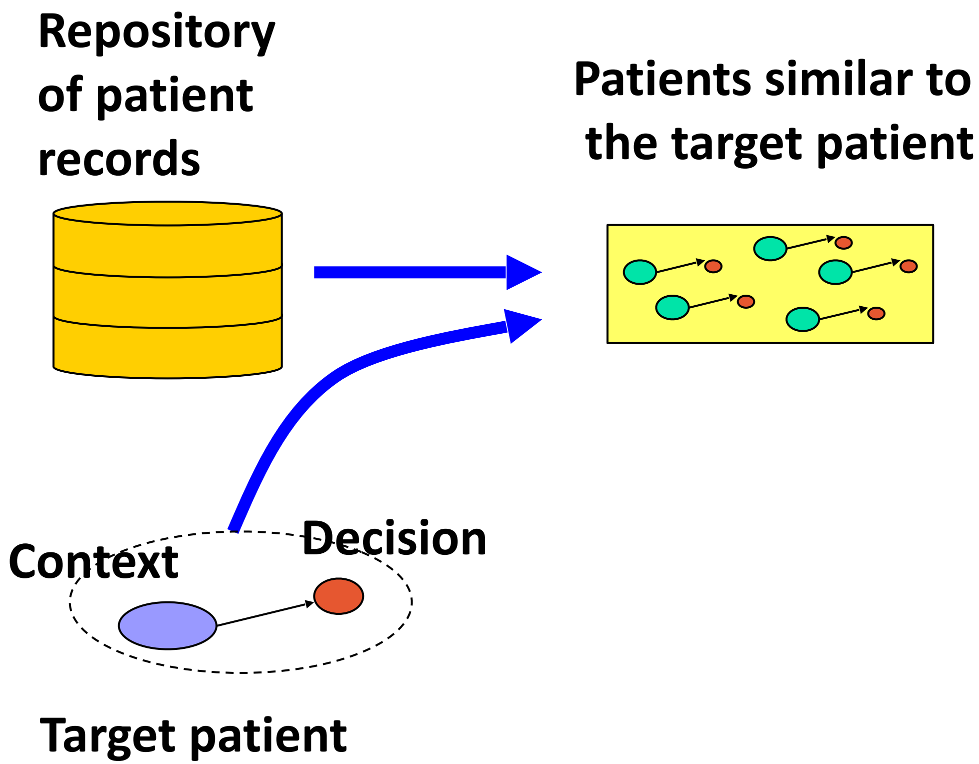
Conditional anomaly detection (CAD)



CAD: subpopulation selection

Step 1. Selection of a relevant patient subpopulation

- **Goal:** identify patients similar to the target patient with respect to the decision variable of interest



CAD: subpopulation selection

Methods:

- **Exact match** on all patient's attributes
  - May not yield a (statistically) sufficient population
- **All patients**
  - Population may be biased by the prior
- **Similarity-based methods**
  - Euclidean metric
    - Ignores feature correlates (double counting)
  - Mahalanobis metrics
    - Features unimportant for the decision are counted
  - Weighted generalized distance metrics
    - Accounts for the importance of the decision

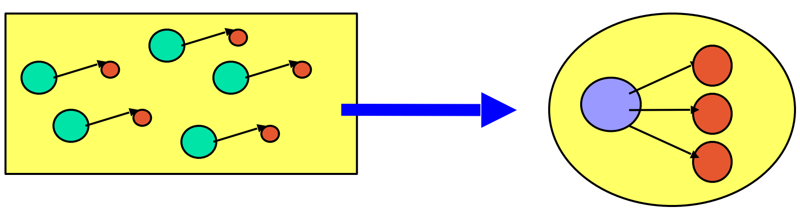
CAD: model building

Step 2. Building a predictive model

- **Goal:** build a probabilistic model that captures how likely individual decision choices are

Patients similar to the target patient

Predictive model for the decision variable



- A task very similar to the classification learning problem but we are trying to learn the predictive distribution

CAD: model building

Step 2. Building a predictive model

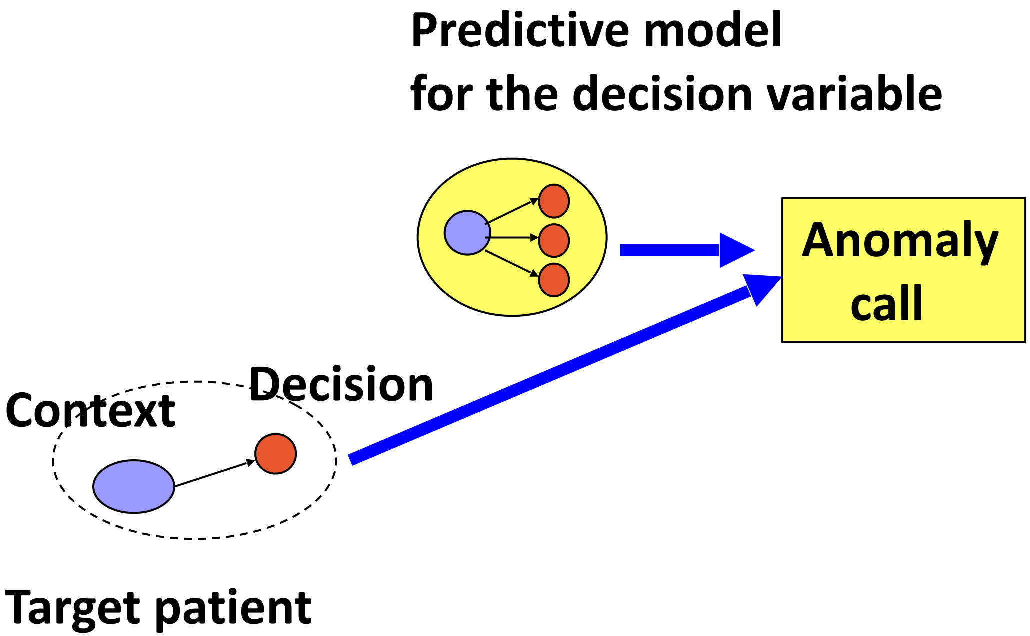
- **Probabilistic models:**
  - Naïve Bayes model
  - Linear discriminant analysis
  - Logistic regression
  - Bayesian belief networks

- Step 1 and step 2 combined is similar to the instance-based learning where a patient-specific classifier is built

CAD: anomaly detection

Step 3. Anomaly (calling) detection

- **Goal:** Given a probabilistic model, and the target patient, decide if the decision made for the target patient is anomalous



CAD: anomaly detection

Step 3. Anomaly detection

- **Approach:** The anomaly call is made with the help of a probability threshold α
  - Absolute criterion: $P(y|\mathbf{x}, \mathbf{M}) < \alpha$
- given the predictive model calculate the probability of observing the decision for the current patient
- if the probability falls bellow the threshold α than report an anomaly

Experimental evaluation

PORT dataset: patients diagnosed with the community acquired pneumonia (2287 cases, 19 attributes)

Decision attribute:
Hospitalization

Demographic factors
Age > 50
Gender (male, female)

Co-existing illnesses:
Congestive heart failure
Cerebrovascular disease
Neoplastic disease
Renal disease
Liver disease

Physical-examination:
Pulse $\geq$125 per min
Respiratory rate $\geq$ 30 per min
Sys. blood pressure < 90mm Hg
Temperature < 35C or > 40C

Lab. & radiograph. findings:
Blood urea nitrogen mg/dl
Glucose mg / dl
Hematocrit < 30 %
Sodium < 130 mmol / l
Art. O2. pressure < 60 mm Hg
Arterial pH < 7.35
Pleural effusion

Experimental evaluation

100 patients selected for the evaluation:

- 21 patients who were found to be anomalous by a simple Naïve Bayes detector
- 79 patients selected randomly

Selected patients presented to a panel of three physicians:

- agreement with the decision
- unsure about the decision
- disagreement with the decision

The case was labeled as "anomalous" if:

- two experts on the panel disagreed with the decision
- or all three were unsure about the decision

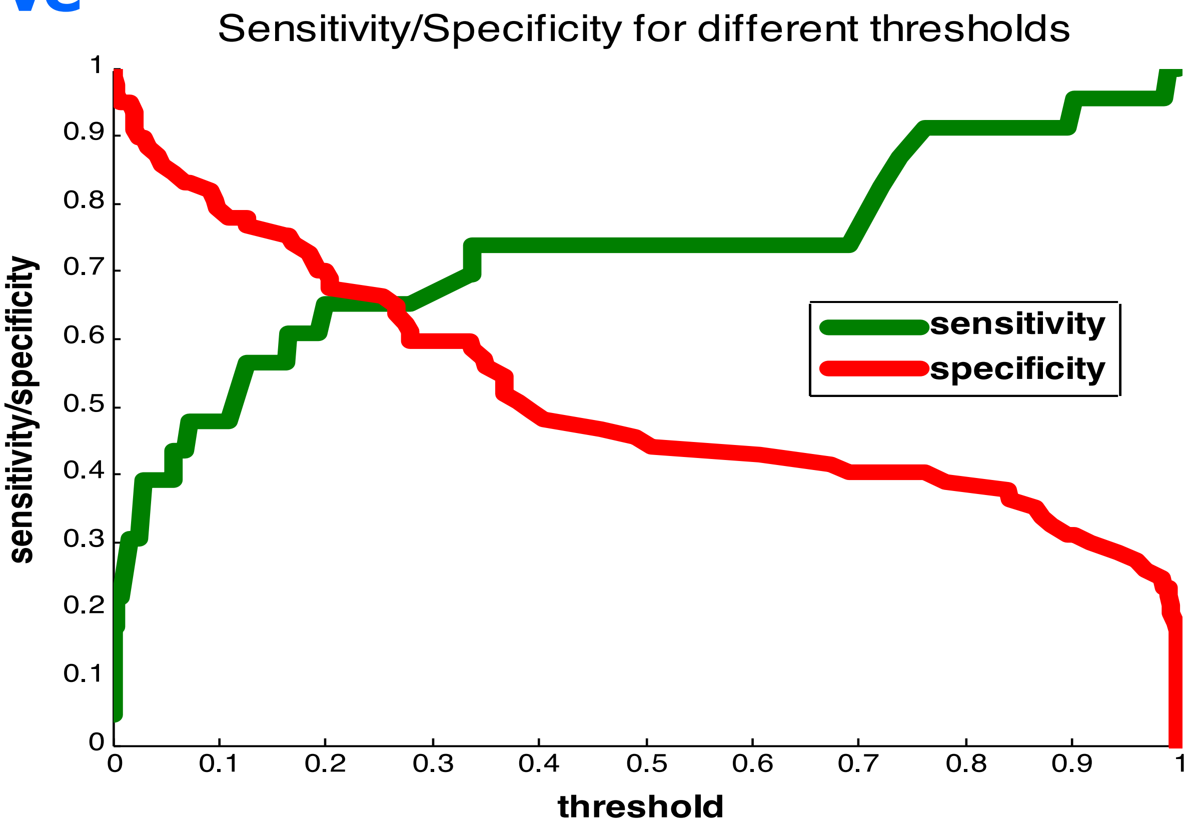
Experimental evaluation: comparison

- **Step 1. Three subpopulation selection methods:**
  - All cases
  - Top 40 Mahalanobis matches
  - Top 40 weighted Mahalanobis matches
    - weights based on the univariate AUC
- **Step 2. Two predictive models:**
  - A naïve Bayes classifier
  - A Bayesian belief network
- **Step 3:**
  - Varied detection thresholds

- <u>**Evaluation metrics:**</u>
  - Area under the ROC curve
  - Area under the Precision-Recall curve

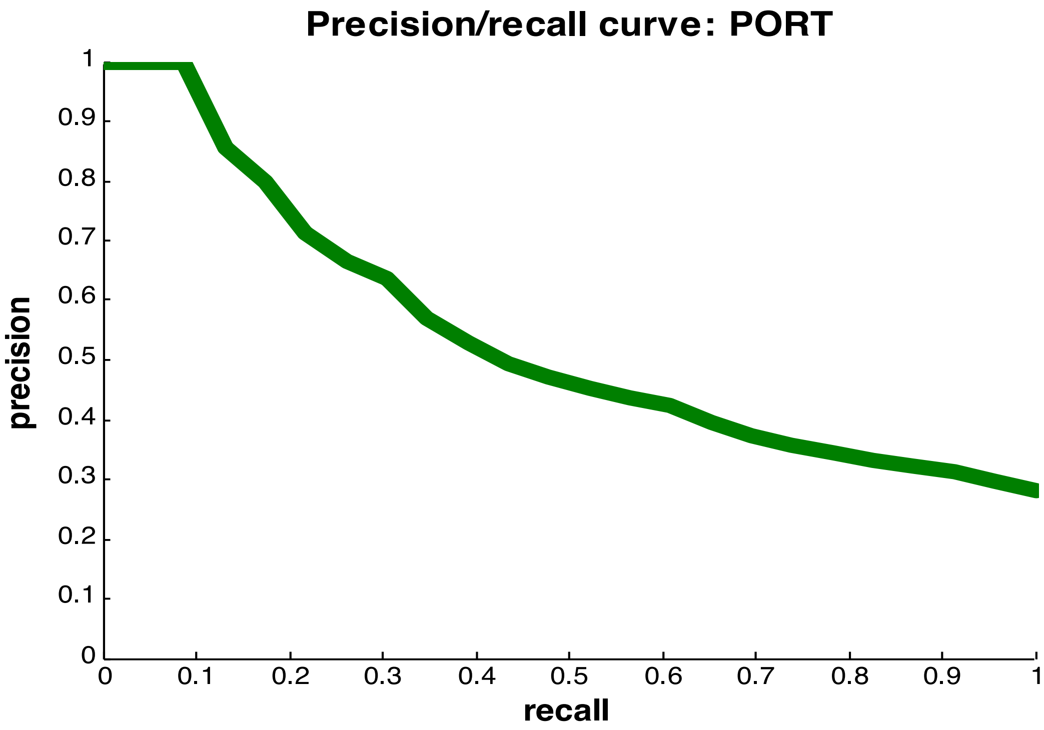
Experimental evaluation: comparison

- Varied detection thresholds lead to different sensitivity-specificity tradeoffs; used to calculate **area under ROC curve**



Experimental evaluation: comparison

- Varied detection thresholds were used to construct the **Precision-Recall curve** (used in the information retrieval)



Experimental evaluation: results

Prob. model	Population selection method	AUC PR	AUC ROC
NB	All cases	0.42	0.74
NB	Mahalanobis	0.49	0.77
NB	Weighted Mahalanobis	0.60	0.80
BBN	All cases	0.56	0.81
BBN	Mahalanobis	0.52	0.79
BBN	Weighted Mahalanobis	0.56	0.80

Experimental evaluation: results

Prob. model	Population selection method	AUC PR	AUC ROC
NB	All cases	0.42	0.74
NB	Mahalanobis	0.49	0.77
NB	Weighted Mahalanobis	0.60	0.80
BBN	All cases	0.56	0.81
BBN	Mahalanobis	0.52	0.79
BBN	Weighted Mahalanobis	0.56	0.80

Naïve Bayes model: Subpopulation selection helps if the predictive model is simpler

Experimental evaluation: results

Prob. model	Population selection method	AUC PR	AUC ROC
NB	All cases	0.42	0.74
NB	Mahalanobis	0.49	0.77
NB	Weighted Mahalanobis	0.60	0.80
BBN	All cases	0.56	0.81
BBN	Mahalanobis	0.52	0.79
BBN	Weighted Mahalanobis	0.56	0.80

BBN is able to capture better (in)dependencies among features, leads to better detection performance than for the Naïve Bayes

Experimental evaluation: results

Prob. model	Population selection method	AUC PR	AUC ROC
NB	All cases	0.42	0.74
NB	Mahalanobis	0.49	0.77
NB	Weighted Mahalanobis	0.60	0.80
BBN	All cases	0.56	0.81
BBN	Mahalanobis	0.52	0.79
BBN	Weighted Mahalanobis	0.56	0.80

BBN benefits less from the subpopulation selection.

- similarity-based subpopulations of size 40
- sample size too small to learn the parameters of a BBN (??)

Results: analysis

- PR curves and statistics indicate:
  - the precision is 0.5
    (1 correct anomaly in 2 every anomalies)
    at 0.53 recall
    (1 out of two anomalies called)
- A biased test dataset (more anomalies).
- A conservative estimate based on the random subset:
  - the precision is 0.32
    (~1 correct anomaly in every 3 anomalies)
    at 0. 3 recall
    (~1 out of three anomalies called)

Results: analysis

Analysis of false negatives

- The evaluators saw only variables the system was able to see
- Many missing attributes
- Disagreements:
  - physicians did not see the reason to hospitalize the patient
  - Causes are among missing variables
  - Our system saw examples of similar cases and did not call the anomaly

Conclusions

- Detection of unusual patient-management decisions by comparison to decisions made for similar patients
- Useful for detection possible patient-management errors
  - The system does not aspire to catch all error decisions but some of them
  - Does it by relying only on past data
- Promising results on pneumonia PORT dataset

Current and Future work

- Application to detection of unusual decisions for post-surgical cardiac patients (~ 5000 medical records)
- Refinement of the conditional anomaly framework
  - Metric learning methods
  - Non-linear similarity metrics
- From detection of possible management errors to the detection of unusual conditions, outcomes, behaviors
  - Design of autonomous agents searching for atypical outcome patterns in large medical record databases. May lead to discovery of subgroups of patients with alternative treatments

Acknowledgments

People who contributed to the project:

- Melissa Saul
- Amy Seibert
- Tomas Singliar
- James Harrison
- Andrew Post

- Funded by NLM: Grant R21-LM009102

Additional slides

CAD: subpopulation selection

Caveats of similarity-based subpopulation selection

- **How to determine the size of the relevant subpopulation?**
  - top k patient matches
  - drop in the distance
  - or weight patients proportionally to their distance
- **Is the population sufficient to make a sound anomaly call?**
  - Exclude outliers based on the average subpopulation distance.

Are “unusual” decisions equal to mistakes

- Some unusual decisions are not equal to errors
  - Differences in management choices across hospitals, or for different physicians may exist.
  - Our system may alert to these decisions.
- **Two solutions:**
  - Flag the alert so that the mistake is not made again.
  - **Abstractions.** Filter out some false positives by incorporating prior knowledge that lets us assess the “equivalence” of some decisions (say two different drugs are equivalent in terms of the goals)

End