Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

Predictive Modelling for Clinical Applications

Harry Freitas da Cruz

Data Management for Digital Health

Summer 2017

Where are we?

Hasso Plattner Institut

Real-world Use Cases

Oncology | Nephrology | Heart Insufficiency | Additional Topics

Data Management & Foundations

Biology Recap

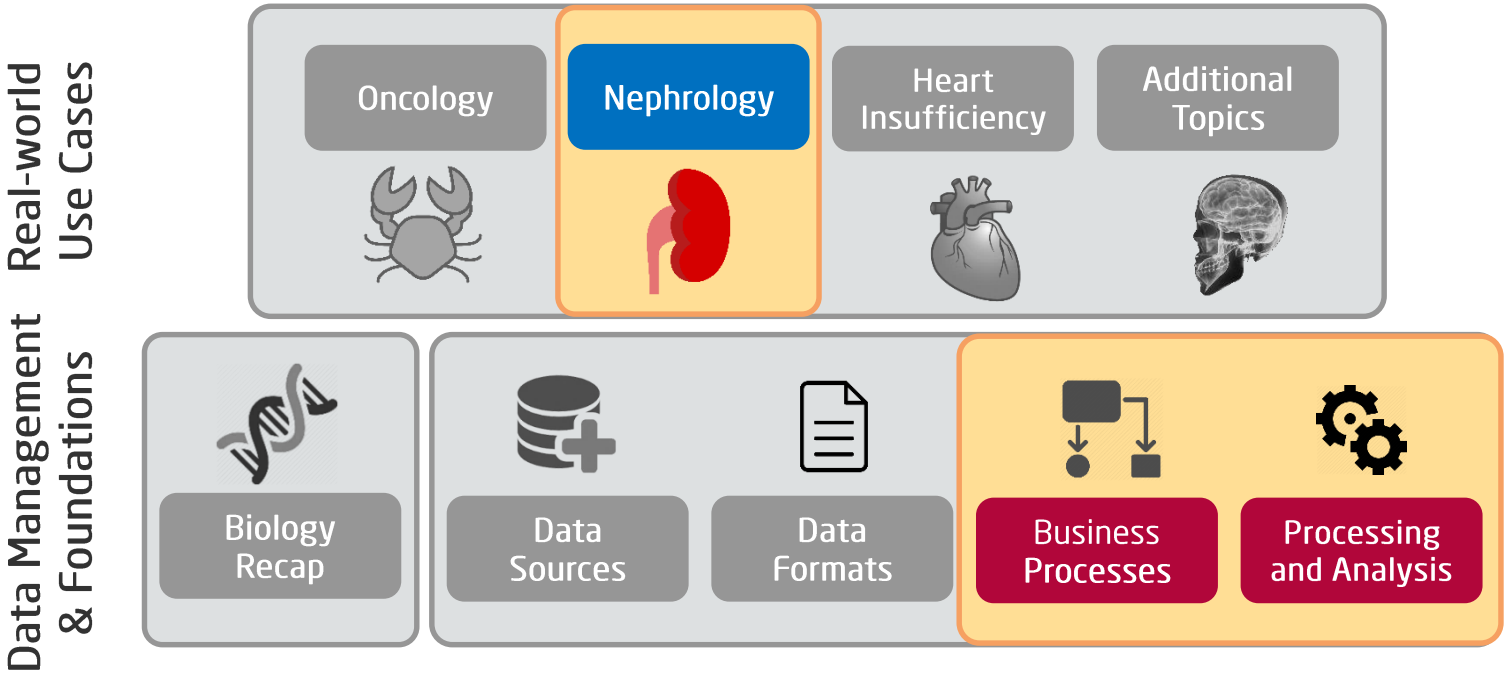
Data Sources

Data Formats

Business Processes

Processing and Analysis

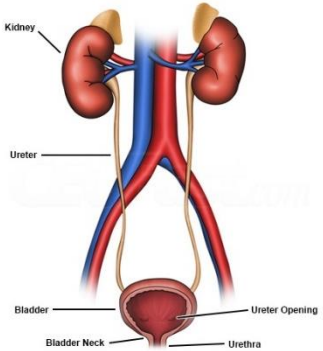
Where are we?


Real-world Use Cases
Oncology
Nephrology
Heart Insufficiency
Additional Topics
Data Management & Foundations
Biology Recap
Data Sources
Data Formats
Business Processes
Processing and Analysis

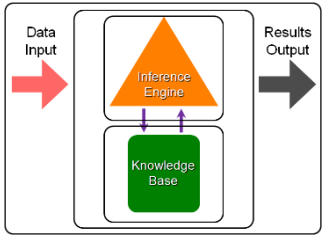
Recap Use Case Nephrology

- The urinary system
- How the kidneys work
- Kidney diseases
- Bayesian networks for AKI
- Machine learning in Nephrology

https://edc2.healthtap.com/ht-staging/user_answer/reference_image/3694/large/Kidney.jpeg



https://ceufast.com/course/urinary-tract-infections-the-unappreciated-giant



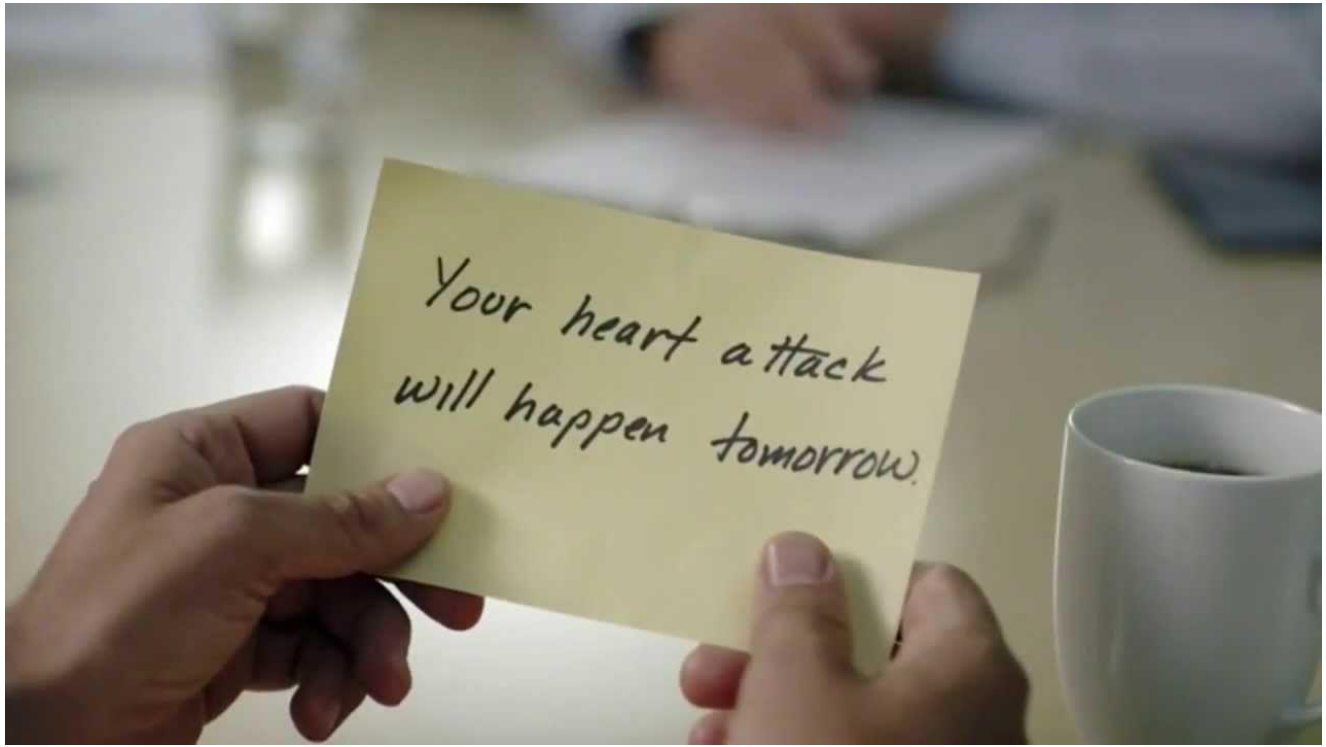
Architecture components of CDSS (Kola, n.d.)

http://sunlightpharmacy.com/wp-content/uploads/2017/03/CKD.jpg

Agenda

- Predictive analytics in healthcare
- Clinical Decision Support Systems (CDSS)
- Clinical Data Repository (CDR)
- Establishing Clinical Prediction Models

Predictive Analytics in Healthcare



https://i.ytimg.com/vi/xEemneA_qwE/maxresdefault.jpg

Predictive Analytics in Healthcare

https://i.ytimg.com/vi/xEemneA_qwE/maxresdefault.jpg

Predictive Analytics in Healthcare
Opportunities and Challenges

Hasso Plattner Institut

Patient safety

Precision medicine

Risk stratification

OPPORTUNITIES

Physician and patient trust

Data standardization

Novel algorithms and tools

CHALLENGES

Predictive Modelling for Clinical Applications

Data Management for Digital Health, Summer 2017

https://www.healthcatalyst.com/three-approaches-to-predictive-analytics-in-healthcare

Predictive Analytics in Healthcare
Clinical Decision Making

- Huge variety of clinical (increasingly genomic) data
- Questions to be answered[1]:
  - What disease does this patient have?
  - Should this patient be treated?
  - Should testing be done?
- Accurate, complete, relevant data
- Reliance on pattern recognition and customary practices
- Evidence-based medicine (clinical guidelines)
- Medical errors are frequent

http://www.mii.ucla.edu/images/research/areas/clinical_decision.png

[1] http://www.merckmanuals.com/professional/special-subjects/clinical-decision-making/introduction-to-clinical-decision-making

Clinical Decision Support Systems Definition

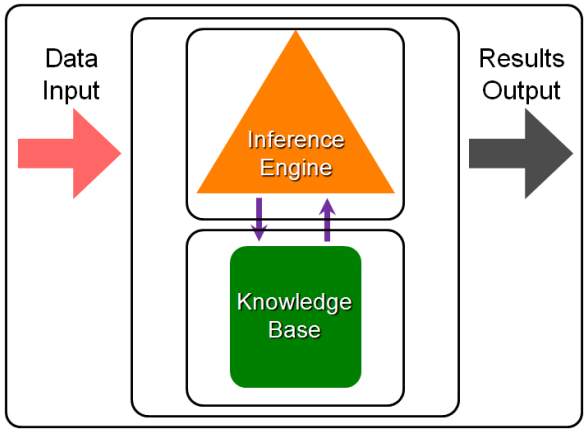
- According to Kabari et al.:

> "Clinical decision support systems (CDSS) provide clinicians, staff, patients, and other individuals with **knowledge** and **person-specific information**, intelligently filtered and presented at **appropriate times**, to enhance health and health care"[1]

[1] L. G. Kabari and E. O. Nwachukwu (2012). Neural Networks and Decision Trees For Eye Diseases Diagnosis, Advances in Expert Systems, Prof. Petrică Vizureanu (Ed.), InTech, DOI: 10.5772/51380.

Clinical Decision Support Systems
Typical Application Scenarios

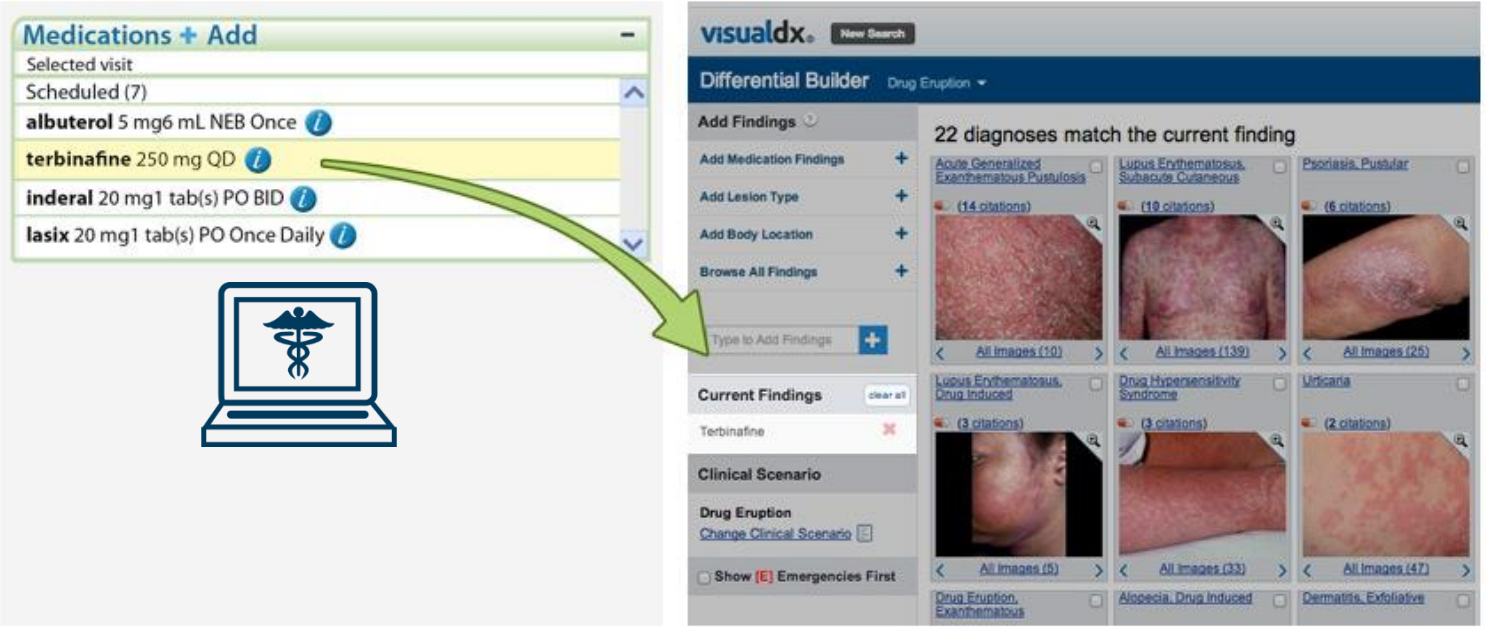
- Usual scenarios
  - Diagnostic support
  - Preventive care
  - Treatment planning / recommendations
- How can this be achieved?
  - Contextual retrieval of highly relevant information
  - Patient-specific reminders and recommendations
  - Organization and presentation of information
- Information logistics / 5 „rights“
  - Information, person, format, channel, time



Architecture components of CDSS (Kola, n.d.)

Berner, E. (2009). Clinical decision support systems: state of the art. AHRQ Publication, (9). Retrieved from https://healthit.ahrq.gov/sites/default/files/docs/page/09-0069-EF_1.pdf

Clinical Decision Support Systems
Contextual Retrieval of Highly Relevant Information



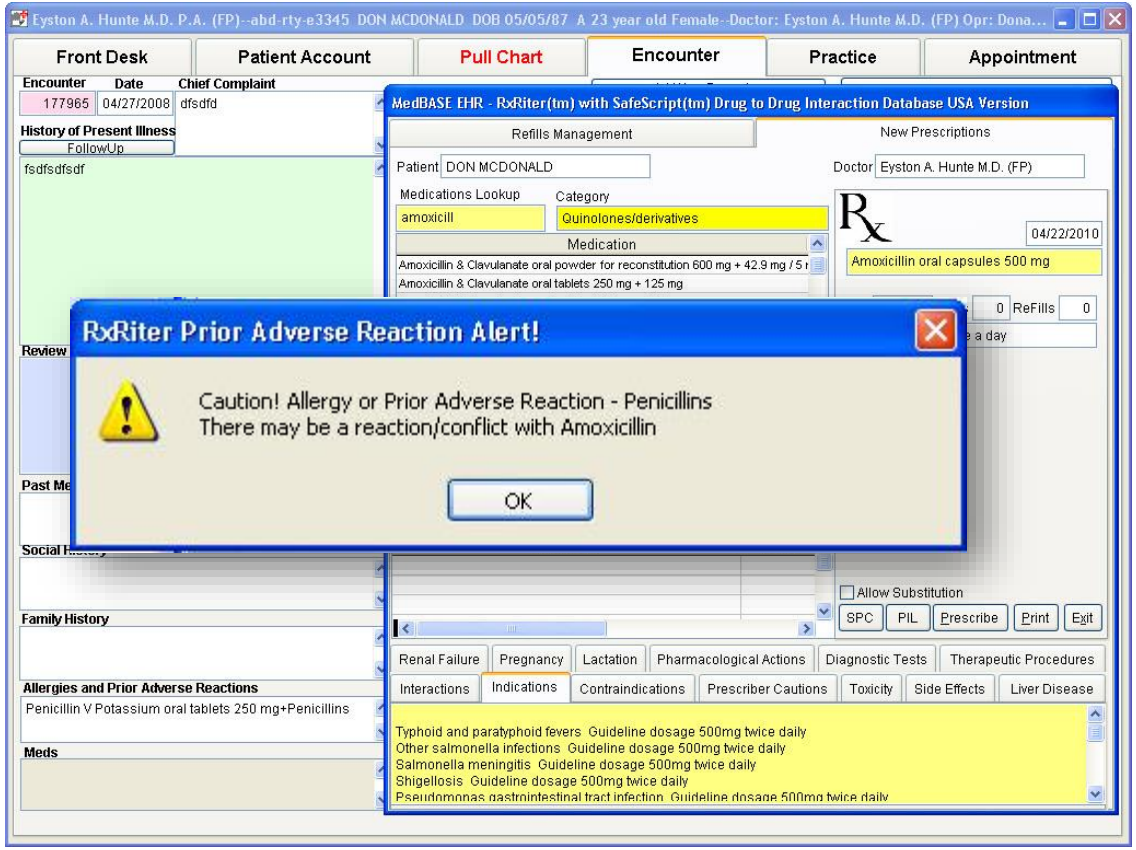
https://www.visualdx.com/benefits/recognize-drug-reactions

Predictive Modelling for Clinical Applications

Data Management for Digital Health, Summer 2017

Clinical Decision Support Systems
Patient-specific Reminders and Recommendations

http://www.ehrscreenshots.com/files/2011/10/rxriter001.jpg

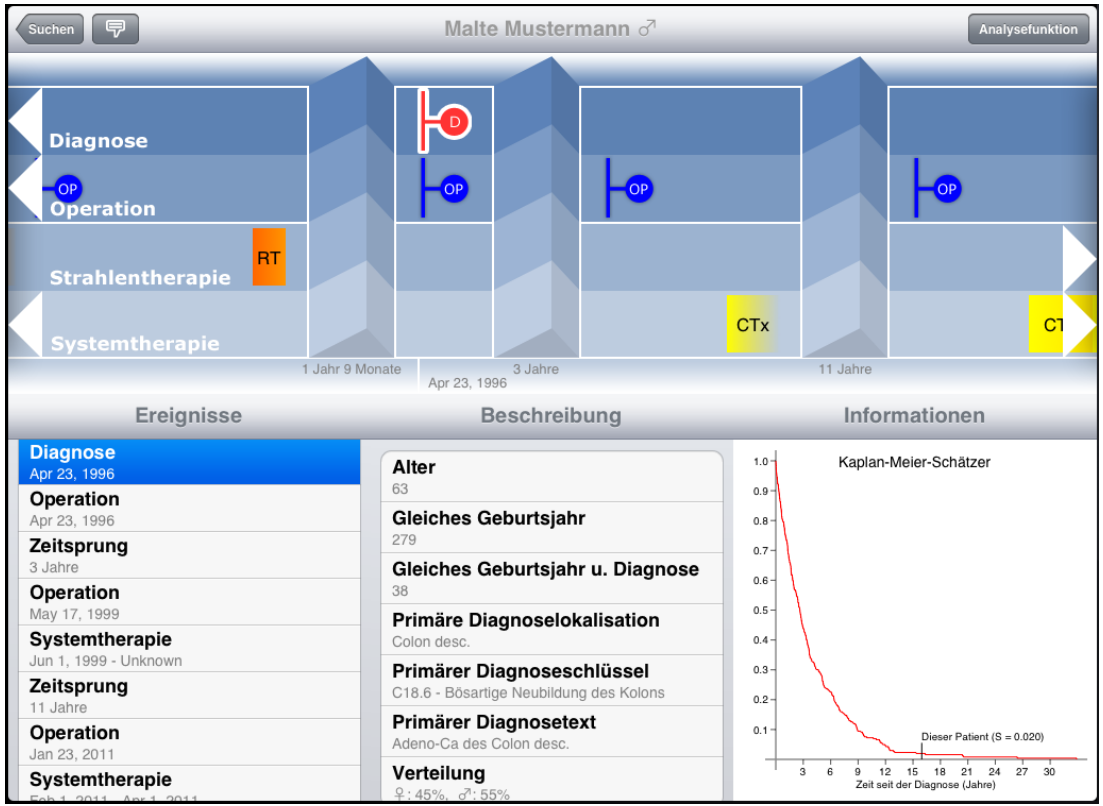


Predictive Modelling for Clinical Applications

Data Management for Digital Health, Summer 2017

Clinical Decision Support Systems
Organization and Presentation of Information

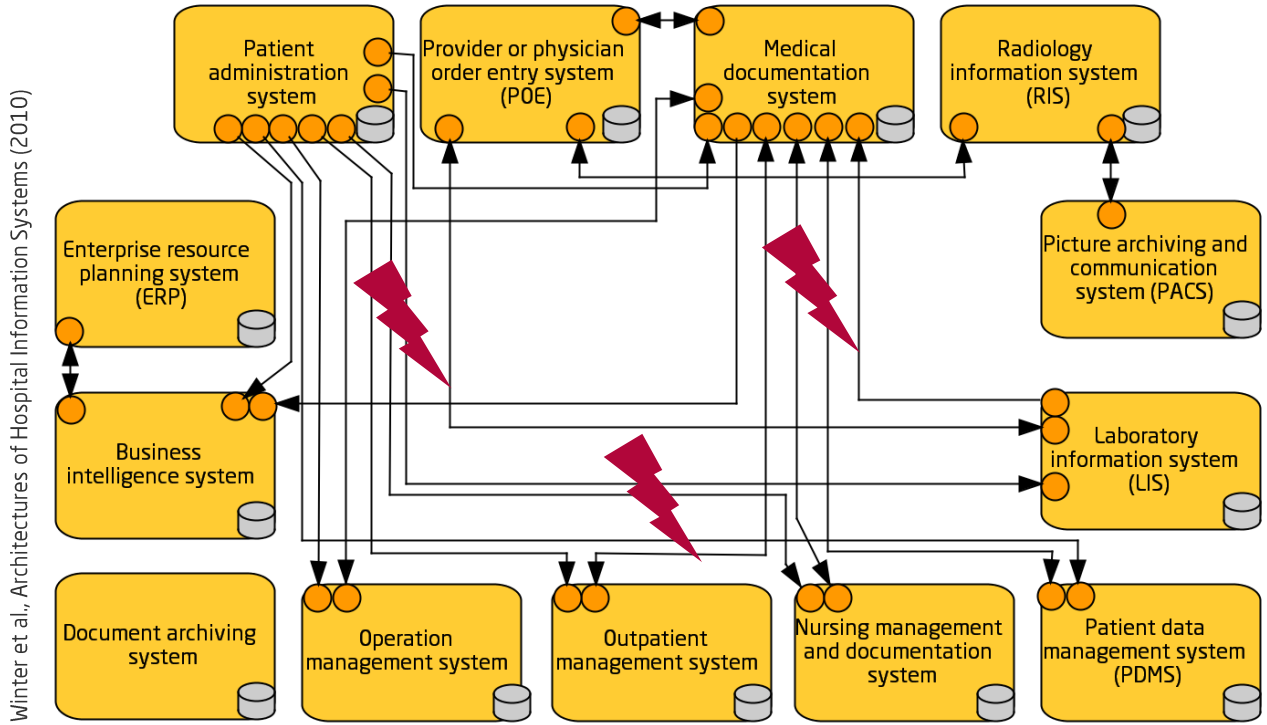
HANA Oncolyzer



Predictive Modelling for Clinical Applications

Data Management for Digital Health, Summer 2017

Clinical Data Repository
Heterogeneous Landscape



Predictive Modelling for Clinical Applications

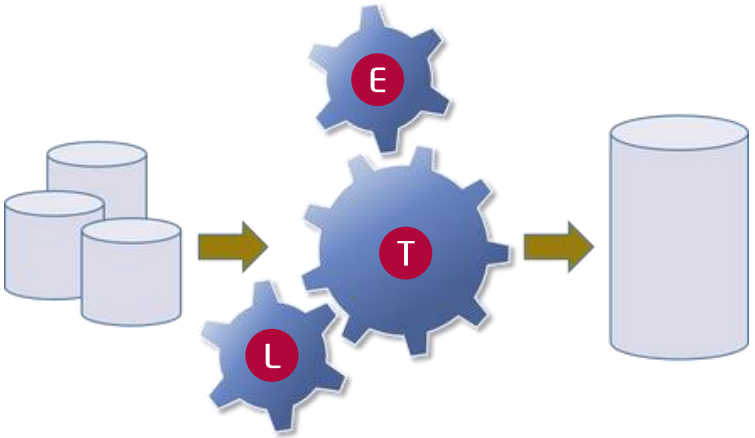
Data Management for Digital Health, Summer 2017

A. Winter, R. Haux, E. Ammenwerth, B. Brigl, N. Hellrung, F. Jahn. Health Information Systems: Architectures and Strategies (2010)

Clinical Data Repository
Challenges and opportunites

- Difficult and costly to develop a general purpose CDR
- Integration of different data sources is a daunting task
- Data inconsistency, redundancies are frequent
- Difficult to adapt to changing needs of users

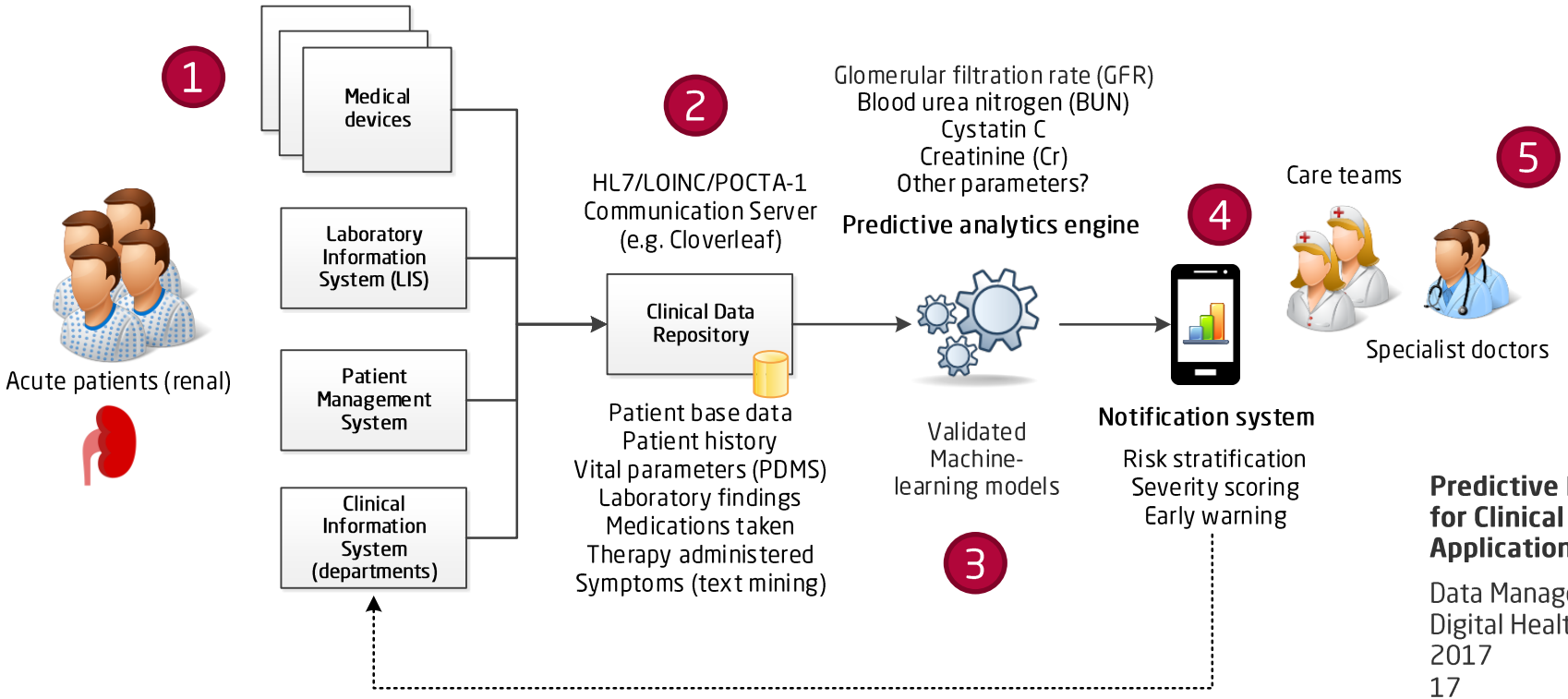
- 360° patient view
- Explore patterns in disease progression and management
- Discover unknown patterns in the data
- Faster hypothesis testing -> Clinical studies



https://discoverbiblog.wordpress.com/2016/11/08/introduction-for-etl/

De Mul, M., Alons, P., van der Velde, P., Konings, I., Bakker, J., & Hazelzet, J. (2012). Development of a clinical data warehouse from an intensive care clinical information system. Computer Methods and Programs in Biomedicine, 105(1), 22–30. http://doi.org/10.1016/j.cmpb.2010.07.002

Clinical Data Repository
Nephrology Use Case (AKI)

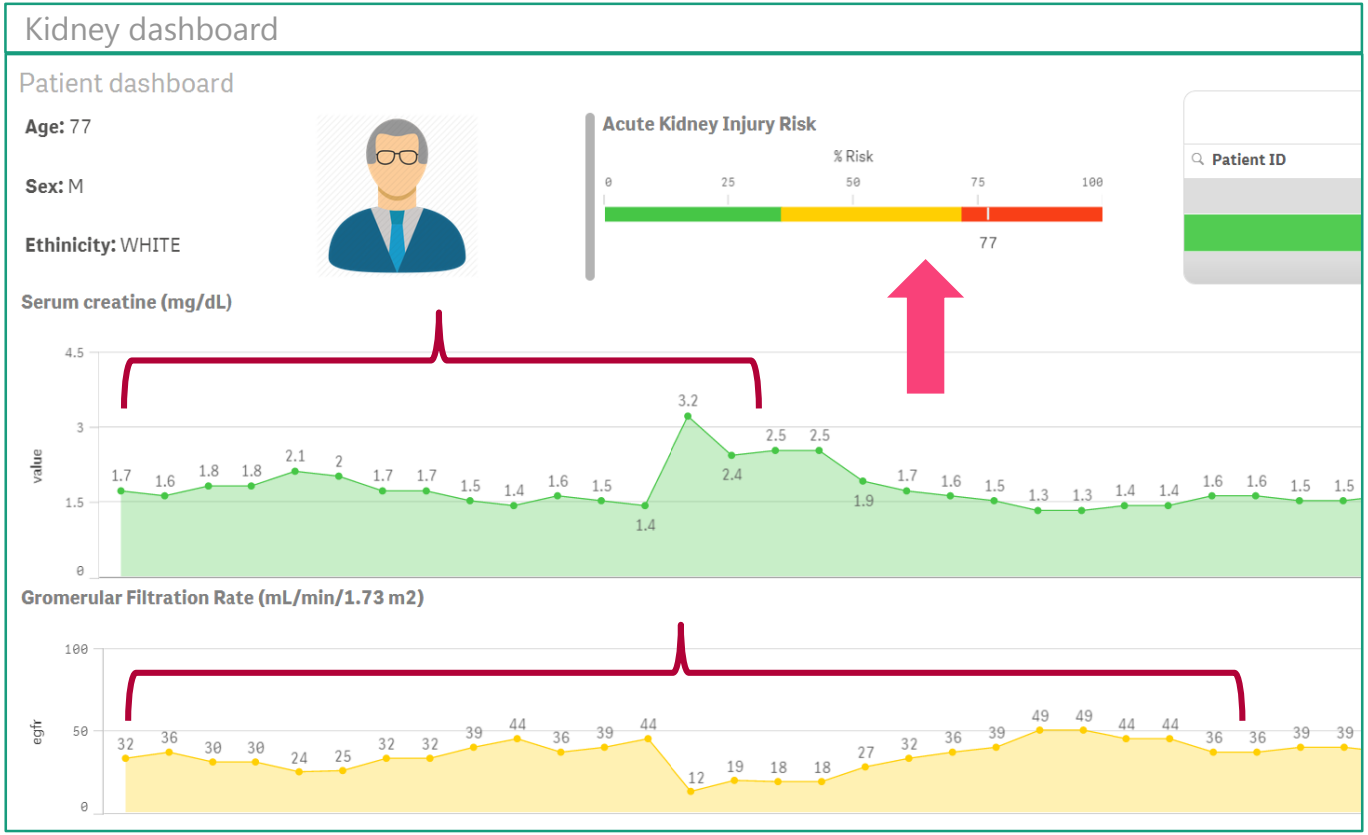


H. F. Cruz, Irish Nephrology Society Annual Scientific Meeting (2016)

Predictive Modelling for Clinical Applications

Data Management for Digital Health, Summer 2017

Clinical Data Repository
Nephrology Use Case (AKI): Patient Dashboard



H. F. Cruz, Irish Nephrology Society Annual Scientific Meeting (2016)

Predictive Modelling for Clinical Applications

Data Management for Digital Health, Summer 2017

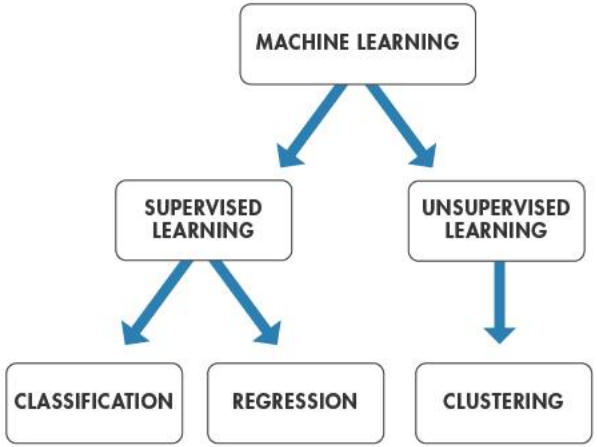
Clinical Prediction Models
Occam's Razor

- Models are an abstraction of reality
- May still be useful depending on the purpose
- Start simple: robust and difficult to break
- As per William of Occam (1287): ontological parsimony
- In diagnostics, not always the case:
  - Hickam's dictum
  - "Patients can have as many diseases as they damn well please"

http://theleanwayconsulting.com/occams-razor-in-business-problem-solving/

Clinical Prediction Models
Supervised or Unsupervised?

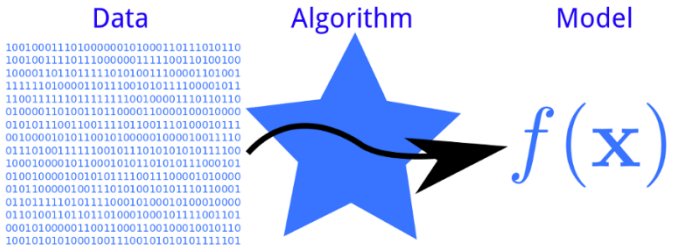
- Supervised learning
  - Labeled data is available
  - Categorical or numerical responses
  - Decision trees, Bayesian nets, ridge regression, etc.
- Unsupervised learning
  - No labeled data
  - Finding hidden patterns in data
  - Hierarchical clustering, k-means, etc.



https://de.mathworks.com/help/stats/machine-learning-in-matlab.html

Clinical Prediction Models
Types of Models

- Classification-based
  - Binary or multi-class
  - Diagnosis, risk stratification, hospital readmission
- Regression-based
  - Optimal drug dosage
  - Treatment plan adjustment
- Survival analysis
  - Time-to-event models
  - Cancer mortality

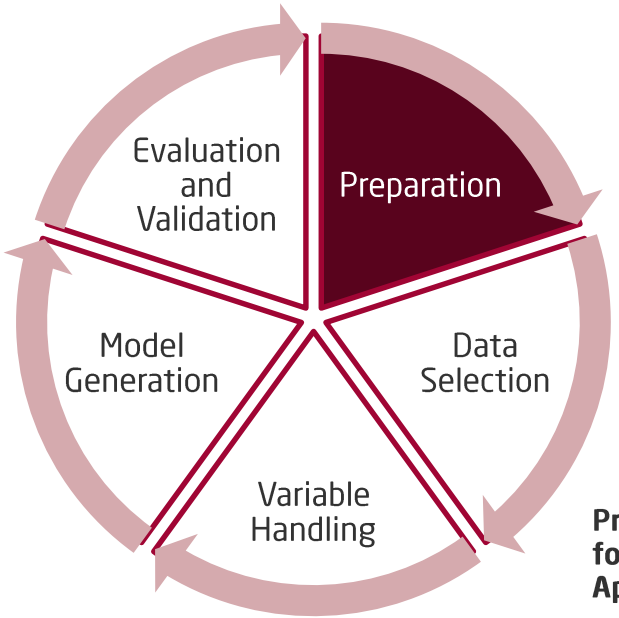


http://phdp.github.io/posts/2013-07-05-dtl.html

Cleophas, T. J., & Zwinderman, A. H. (2013). Machine Learning in Medicine (Vol. 53). Springer Berlin Heidelberg. http://doi.org/10.1017/CBO9781107415324.004

Clinical Prediction Models
Establishing a CPM: Preparation and Dataset Selection

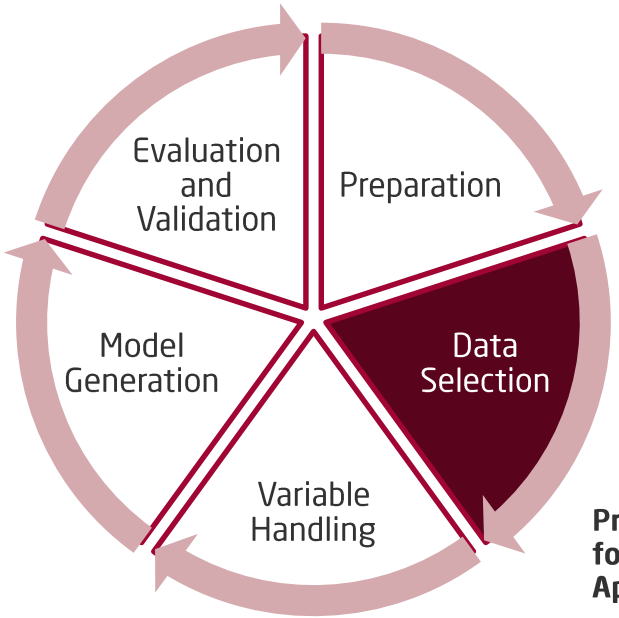
- Step 1: Preparation
  - What is the target outcome?
  - What is the target patient?
  - What is the target user?
- Step 2: Dataset selection
  - Is the data needed available?
  - Is the data representative?
  - What is the validation strategy?



Lee, Y.-H., Bang, H., & Kim, D. J. (2016). How to Establish Clinical Prediction Models. Endocrinology and Metabolism, 31(1), 38. http://doi.org/10.3803/EnM.2016.31.1.38

Clinical Prediction Models
Establishing a CPM: Preparation and Dataset Selection

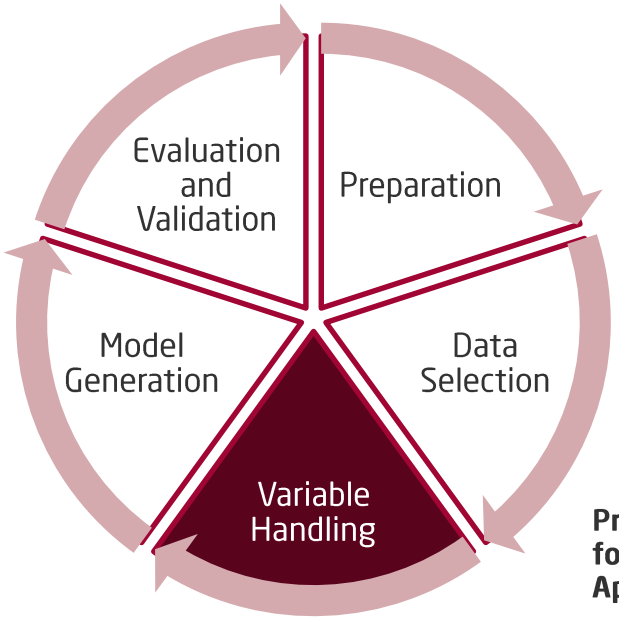
- Step 1: Preparation
  - What is the target outcome?
  - What is the target patient?
  - What is the target user?
- Step 2: Dataset selection
  - Is the data needed available?
  - Is the data representative?
  - What is the validation strategy?



Lee, Y.-H., Bang, H., & Kim, D. J. (2016). How to Establish Clinical Prediction Models. Endocrinology and Metabolism, 31(1), 38. http://doi.org/10.3803/EnM.2016.31.1.38

Clinical Prediction Models
Establishing a CPM: Variable Handling

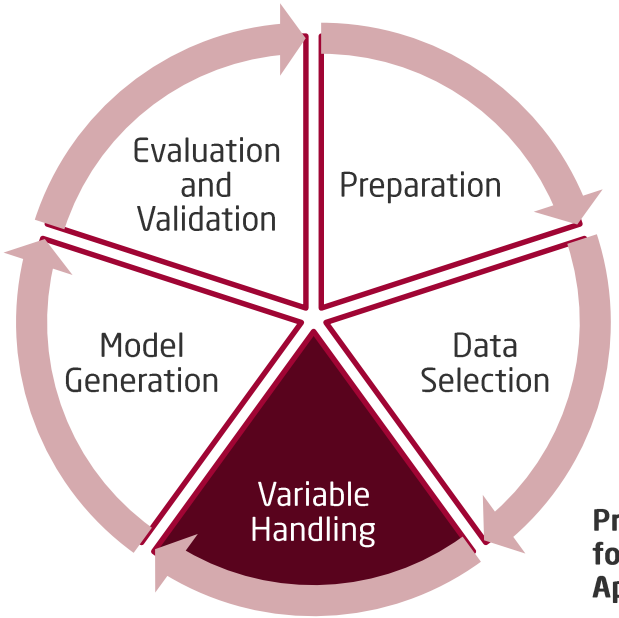
- Expert judgement often necessary
- Previous significant factors should be used
- Avoid predictors that are possible correlated
- Merging categorical variables should be considered
- Scale matters in continuous variables: consistency
- Nominalization often needed for continuous variables
- Consider scale transformation (e.g. log)
- Normalization of values (e.g. from 0-1)
- Identify and handle outliers



Lee, Y.-H., Bang, H., & Kim, D. J. (2016). How to Establish Clinical Prediction Models. Endocrinology and Metabolism, 31(1), 38. http://doi.org/10.3803/EnM.2016.31.1.38

Clinical Prediction Models
Establishing a CPM: Missing Data

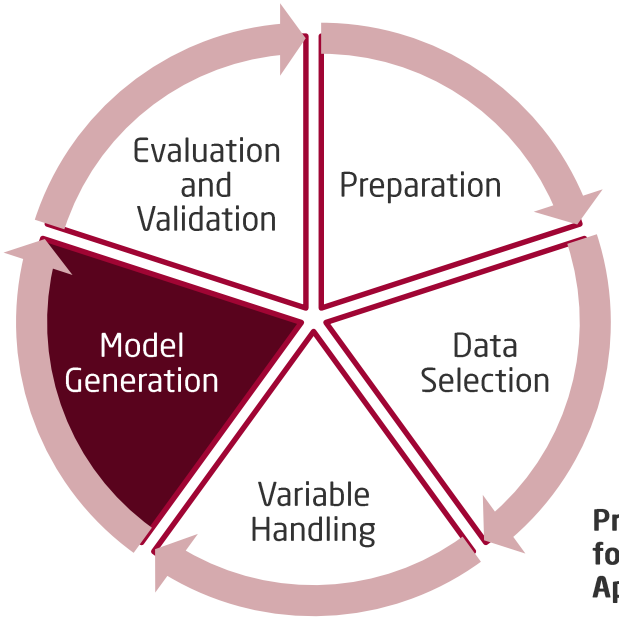
- Leaving out the missing ones: complete case analysis
- Single imputation x multiple imputation
  - Using “other” or “unknown”
  - Averaging occurances, median or mean
- MICE (Multiple Imputation using Chained Equations)
- Regression model from the existing variables



Wells, B. J., Chagin, K. M., Nowacki, A. S., & Kattan, M. W. (2013). Strategies for handling missing data in electronic health record derived data. EGEMS (Washington, DC), 1(3), 1035. http://doi.org/10.13063/2327-9214.1035

Clinical Prediction Models
Establishing a CPM: Model Generation

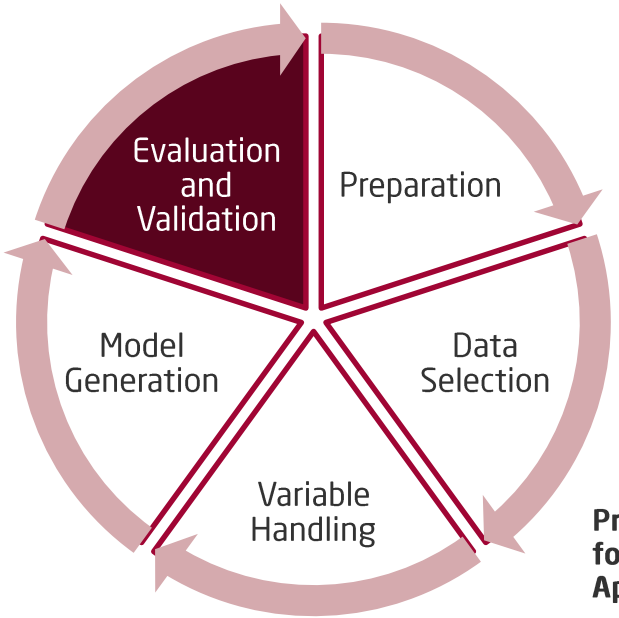
- Selecting the proper algorithm
- Performing feature selection
  - Backward elimination
  - Stepwise selection
- Trade-offs between goodness of fit and complexity
  - Akeike Information Criterion
  - Bayes Information Criterion
- Perform parameter tuning
  - Optimization of hyperparameters



Lee, Y.-H., Bang, H., & Kim, D. J. (2016). How to Establish Clinical Prediction Models. Endocrinology and Metabolism, 31(1), 38. http://doi.org/10.3803/EnM.2016.31.1.38

Clinical Prediction Models
Establishing a CPM: Evaluation and Validation

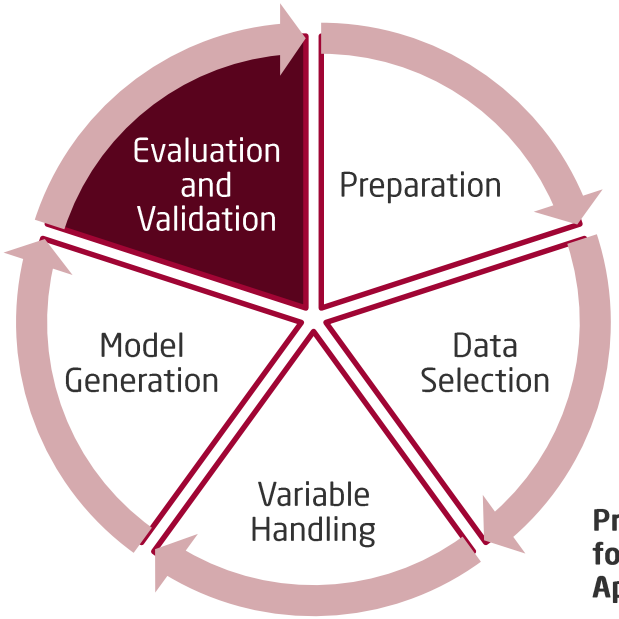
- Internal validation
  - Cross-validation
  - Bootstrapping
- External validation
  - Using a different data source
  - Ensure transportability and generalizability
- Measures of performance
  - ROC Curve
  - R^2, p-values
- True positive rate (TPR) / true negative rate



Clinical Prediction Models
Establishing a CPM: Measures of Performance

Measure
Sensitivity and specificity
Discrimination (ROC/AUC)
Predictive values: positive, negative
Likelihood ratio: positive, negative
Accuracy: Youden index, Brier score
Number needed to treat or screen
Calibration: Calibration plot, Hosmer-Lemeshow test
R^2 statistical significance: p-value (e.g. likelihood ratio test)
Magnitude of association, e.g., β coefficients, odds ratio
Model quality: Akeike IC/ Bayes IC
Net reclassification index and integrated discrimination improvement
Net benefit
Cost-effectiveness

Measures of model performance (Lee 2016)

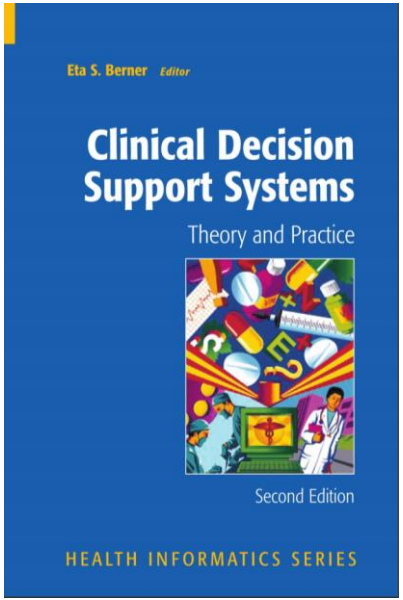


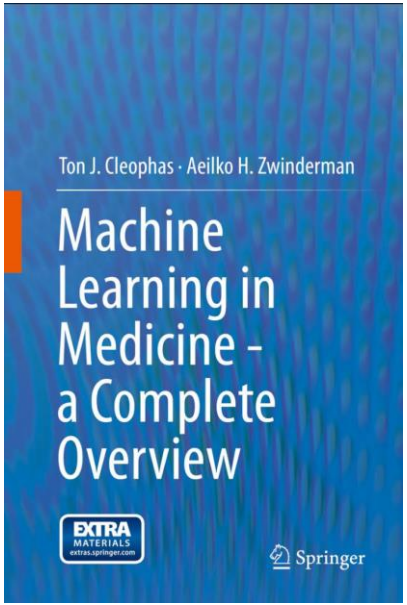
Lee, Y.-H., Bang, H., & Kim, D. J. (2016). How to Establish Clinical Prediction Models. Endocrinology and Metabolism, 31(1), 38. http://doi.org/10.3803/EnM.2016.31.1.38

What to Take Home?

- There is lots of potential for health care analytics
- Clinical Decision Support Systems offer a way to tap into that
- Step-by-step process to establishing clinical prediction models
  - Preparation
  - Data selection
  - Model generation
  - Evaluation and validation

To Know More


Eta S. Berner Editor
Clinical Decision Support Systems
Theory and Practice
Second Edition
HEALTH INFORMATICS SERIES


Ton J. Cleophas · Aeilko H. Zwinderman
Machine Learning in Medicine - a Complete Overview
EXTRA MATERIALS
Springer

What's Coming Next?

- Applying the step-by-step process with an example and tool (Rapid Miner)
- Instructions for the exercise