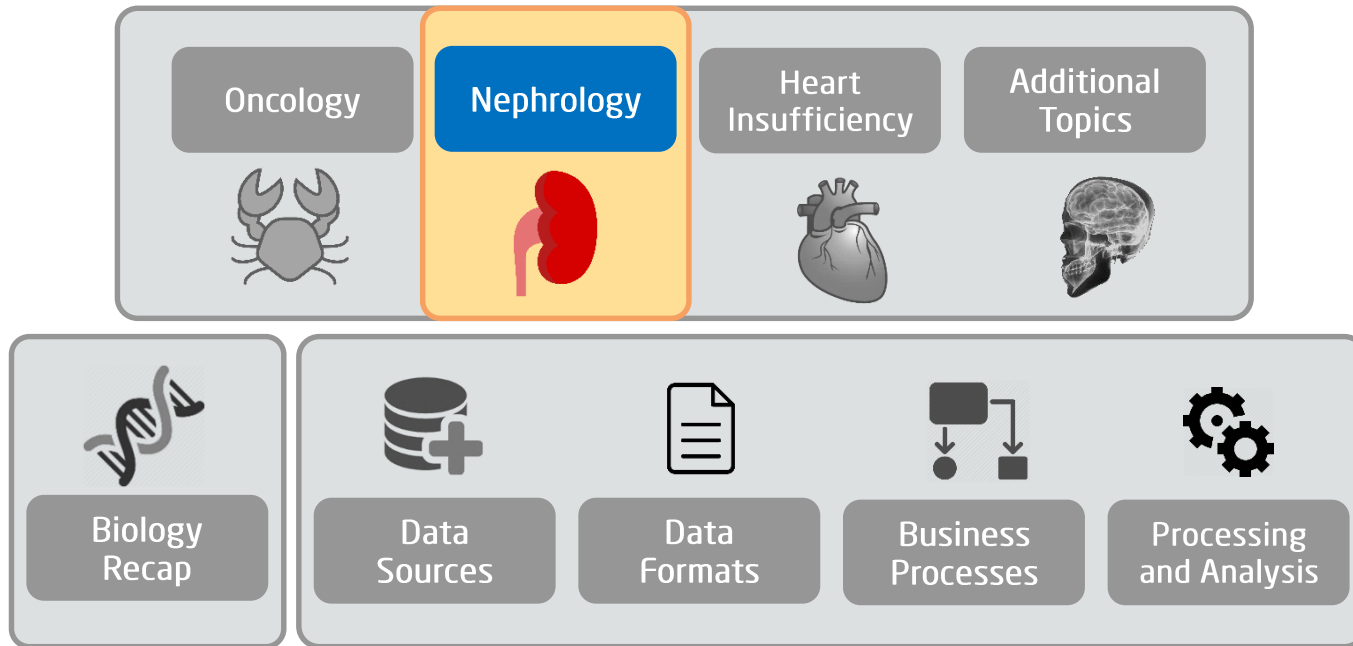# Evaluation Exercise II

Harry Cruz, Milena Kraus

Data Management for Digital Health

Summer 2017

Exercise II

Real-world Use Cases

- Oncology
- Nephrology
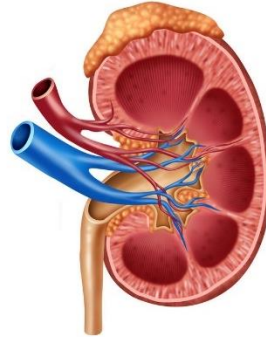- Heart Insufficiency
- Additional Topics

Data Management & Foundations

- Biology Recap
- Data Sources
- Data Formats
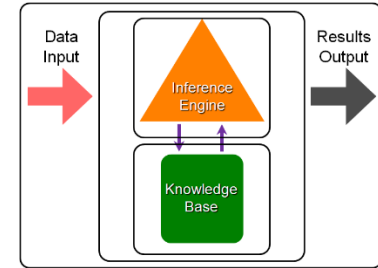- Business Processes
- Processing and Analysis
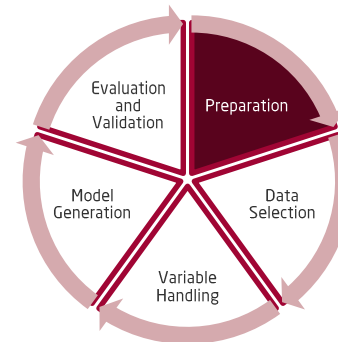
Exercise II
Topics

- Nephrology fundamentals

- Clinical Decision Support Systems

- Clinical Data Repository

- Clinical Prediction Models

- Building a CPM with RapidMiner

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

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

Lee, Y.-H., Bang, H., & Kim, D. J. (2016)

**Walk-through
Exercise II**

Data Management for
Digital Health, Summer
2017

3

Exercise II
Key Stats

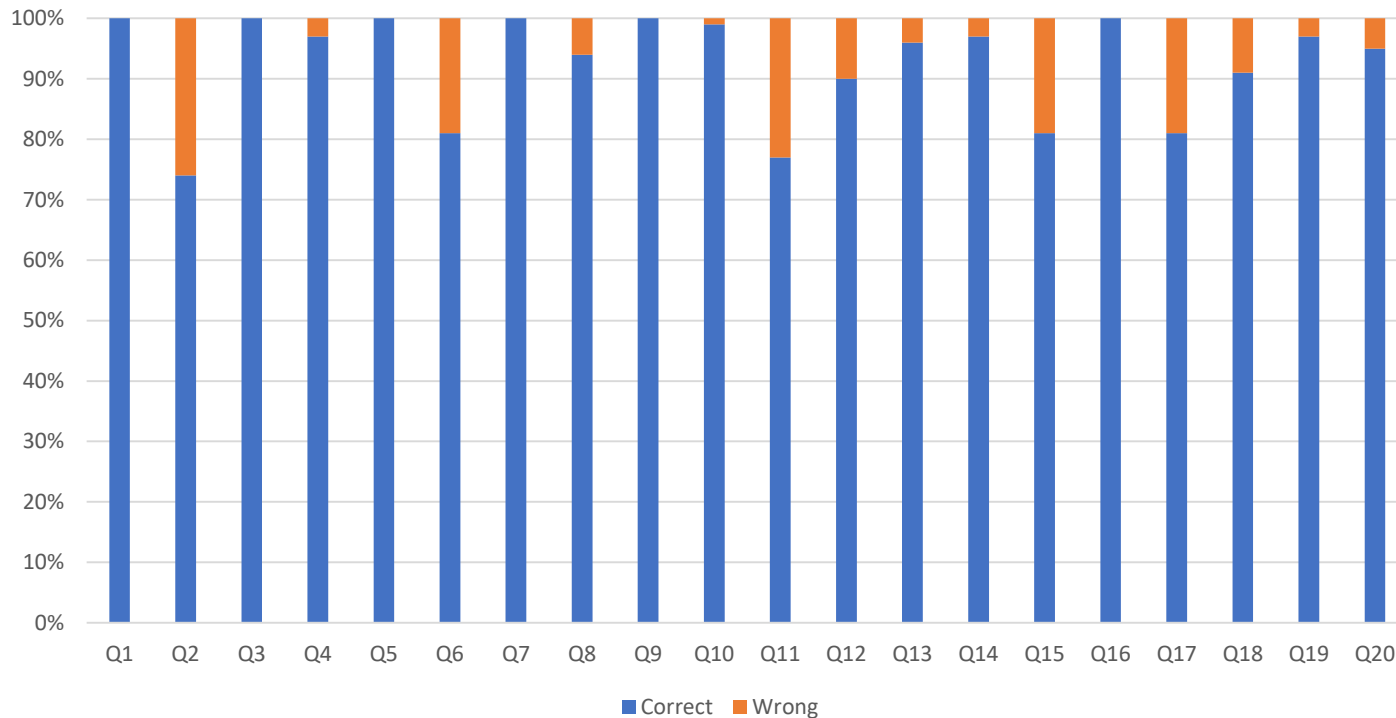| 20 Questions 35 Points | 32 Students 32 Passed | Average score 28.91 / 83% | Average time 105min |

Exercise II
Key Stats

Q2. What are possible reasons for acute kidney disease at the *renal* level? Please choose all that apply:

- **Physical damage to the kidneys**
- Blockage in the urethra
- **Inflammation of the glomeruli, i.e. glomerulitis**
- Bladder cancer

Kidney Disease(s)
Acute Kidney Injury (AKI)

- Sudden and severe drop of renal function

- Increased levels of urea and creatinine

- May or may not be reversible

- Leads to poor patient outcomes

- Affects between 7 and 18% of hospital patients[1]

- Etiology
  - Pre-renal
  - Renal
  - Post-renal

Pre-renal

Renal

Post-renal
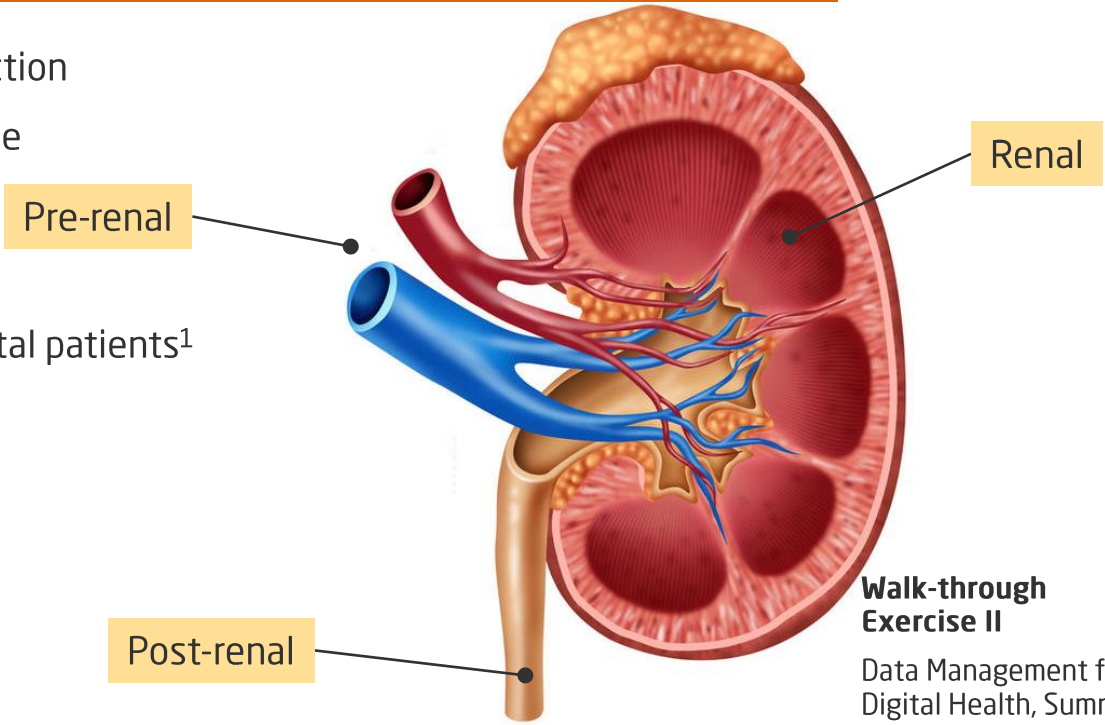
Walk-through Exercise II

Data Management for Digital Health, Summer 2017

7

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

Lewington AJ, Cerdá J, Mehta RL. Raising Awareness of Acute Kidney Injury: A Global Perspective of a Silent Killer. Kidney international. 2013;84(3):457-467. doi:10.1038/ki.2013.153.

Q6. With respect to a Clinical Data Repository (CDR) and its implementation, please choose all correct alternatives:

- The integration of heterogenous data sources into a CDR is often time-consuming, e.g. as it requires the combination and harmonization of different data formats.

- CDRs are planned on top of existing distributed clinical data sources to harmonize data formats and improve data quality.

- Medical researchers usually have clear requirements for CDR that remain stable over time.

- Extract Transform Load (ETL) methods known from business warehouses cannot be used in a clinical context due to privacy regulations.

ETL can also be used in a medical context
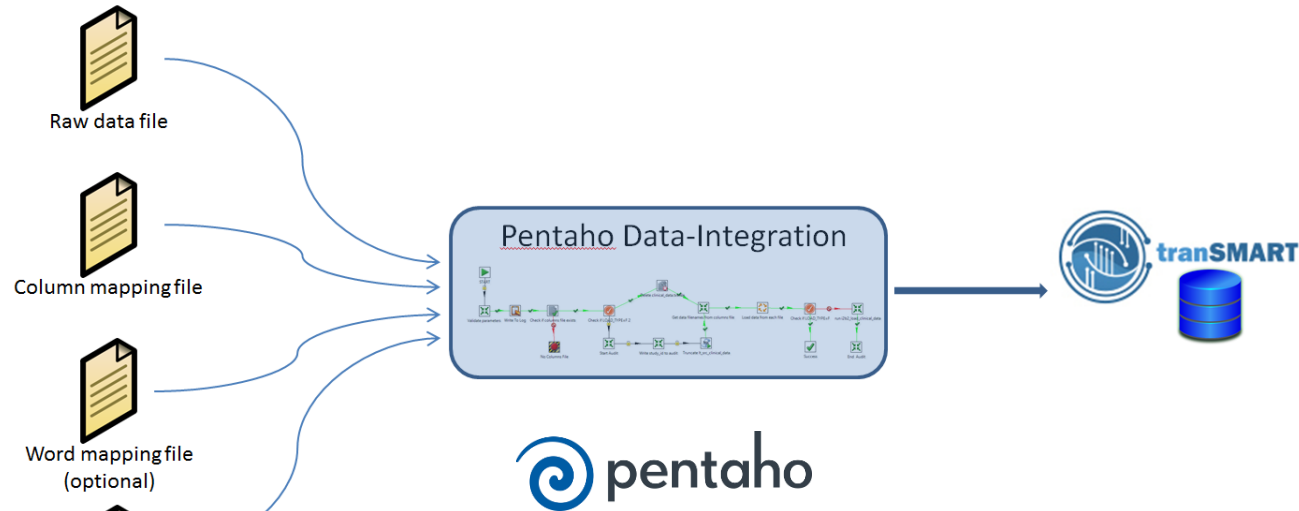
https://wiki.transmartfoundation.org/display/transmartwiki/5+Clinical+Data

Privacy issues concerning medical data (and others)

- EU General Data Protection Regulation

- Unification of privacy regulations in Europe

- Cross-border health research stands to be facilited

- GDPR comes into force in mid 2018

- Secondary use of medical data

- Little impact on biomedical data research

Rumbold, J. M. M., & Pierscionek, B. (2017). The Effect of the General Data Protection Regulation on Medical Research. *Journal of Medical Internet Research*, *19*(2), e47. http://doi.org/10.2196/jmir.7108

Q11. A critical step in the development of a CPM is its validation. Please choose all alternatives that apply:

- A ROC curve is an analysis instrument suitable for comparing the performance of different regression models with each other.

- **In the absence of external validation data sets, performance can be measured by cross-validation using subsets of the training data.**

- **A model presenting a ROC AUC = 0.5 is no better than using a random predictor, i.e. dicing.**

- A complex model is to be preferred over a simpler model since the former usually presents higher precision than the latter.

Q15. A new biomarker-based test to diagnose glomerulitis can correctly classify 80% of sick patients...

- The test presents an accuracy of 0.8 and precision of 0.4.

- **The test has a recall of 80% and a precision of 40%.**

- The test has a higher specificity than sensitivity.

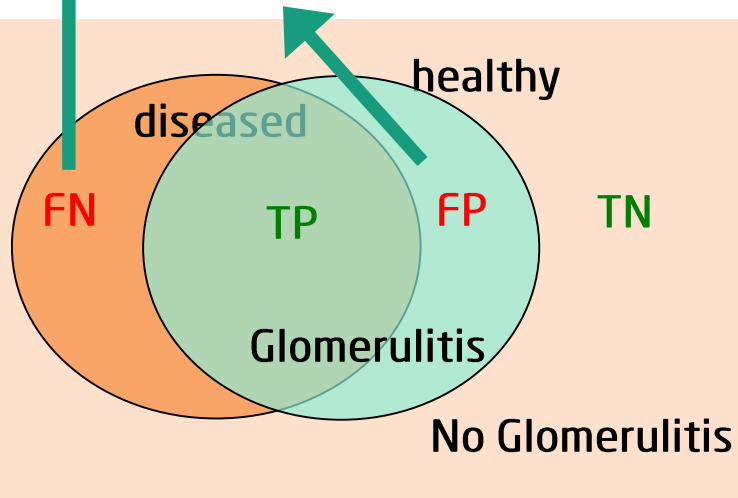- **Specificity and true negative rate of the test is 70%.**

The Biomarker-based Test for Glomerulitis

recall (sensitivity) $= TP / (TP + FN) = 16 / (16 + 4) = 16 / 20 = 0.8$

precision $= TP / (TP + FP) = 16 / (16 + 24) = 16 / 40 = 0.4$

Specificity = TNR = TN / (TN + FP)

TN = 80 (all negatives) – 24 (false positives) = 56

$= 56 / (56 + 24) = 56 / 80 = 0.7$

healthy

diseased

FN   TP   FP   TN

Glomerulitis

No Glomerulitis

specificity $= TN / (TN + FP)$

Q17. Now, we want to examine the performance of the clinical model just created…
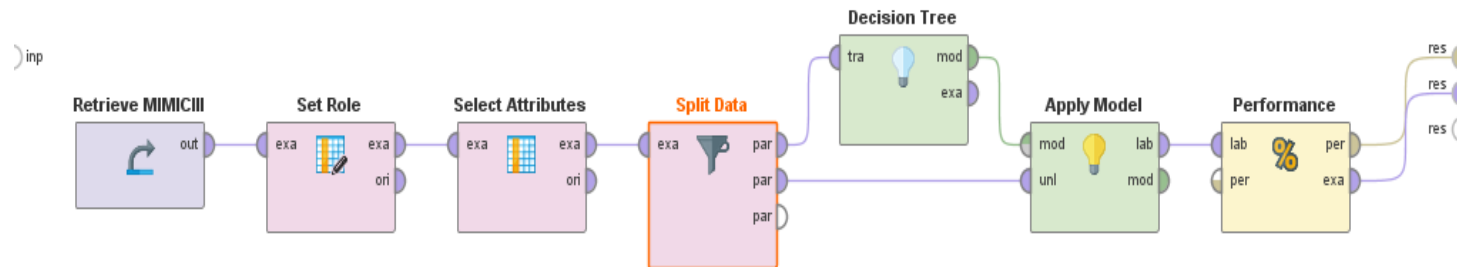
- **The model as such performs overwhelmingly better at classifying patients as not having AKI.**

- **A simple split 50:50 is often not sufficient as validation strategy.**

- Accuracy for AKI=no is around 99,99%

- In total, the model classified 100 surgery patients as having AKI
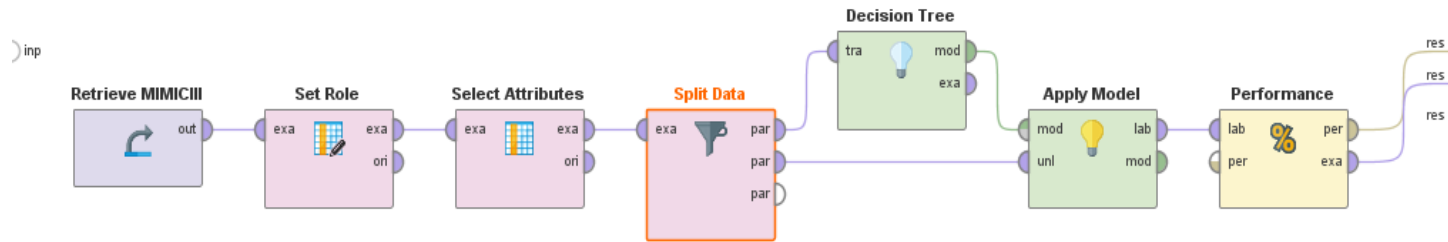
Q17. Now, we want to examine the performance of the clinical model just created…

accuracy: 91.26%

	true no	true yes	class precision
pred. no	5869	531	91.70%
pred. yes	41	100	70.92%
class recall	99.31%	15.85%	