Data, data, everywhere,
Let's all have a drink (*of data*)

Christopher McCudden, PhD, DABCC, FAACC, FCACB

Eastern Ontario Regional Laboratory Association

University of Ottawa

The Ottawa Hospital

Disclosures

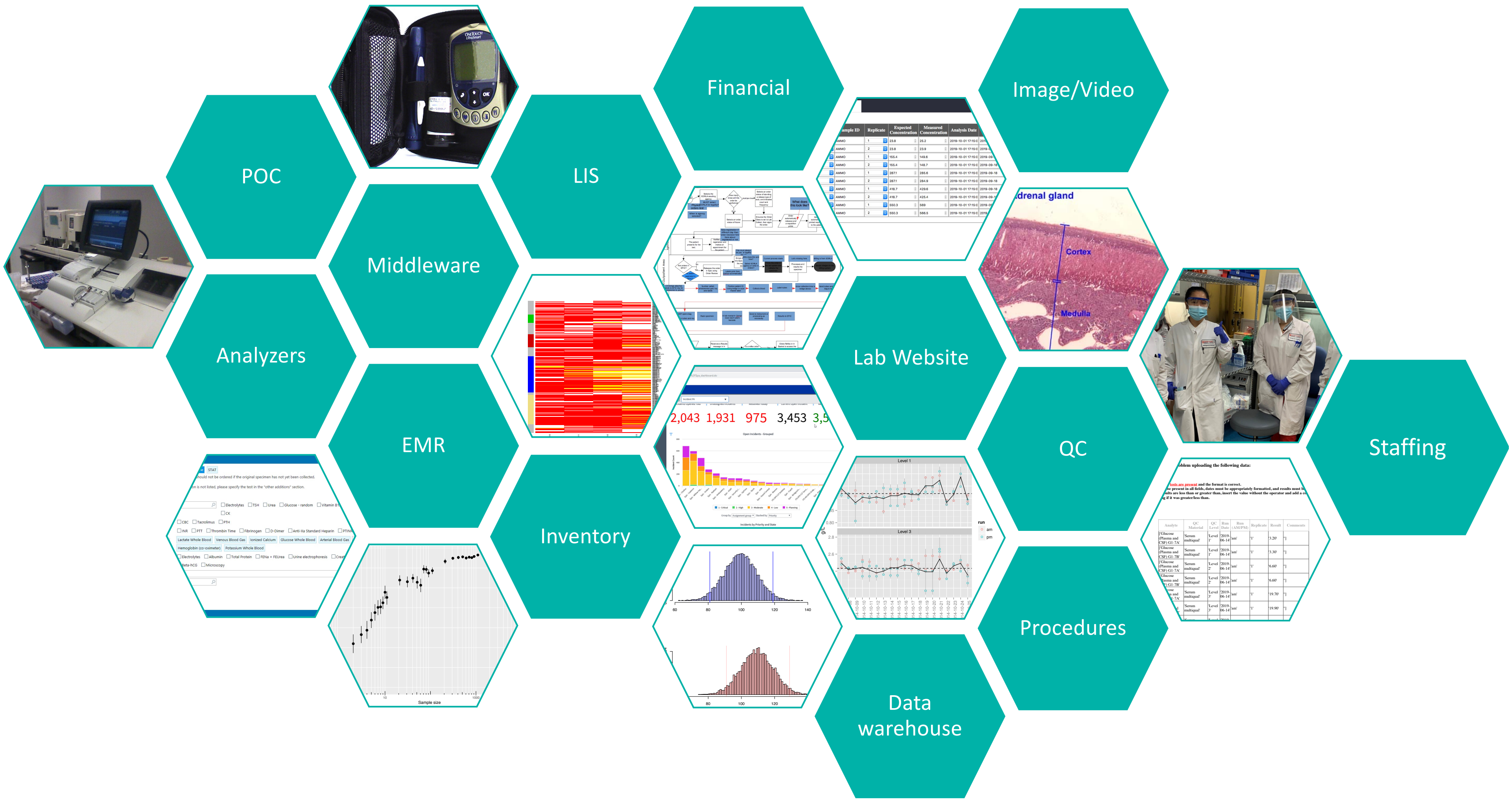
None

Learning Objectives

1) Identify applications and challenges of large datasets
2) Define and differentiate machine learning, statistical learning, and artificial intelligence
3) Identify emerging uses of machine learning in the laboratory
4) Explain how laboratorians can participant in algorithm development and implementation

Part I: Lab Data Opportunities

Lab Data Sources



POLL #1

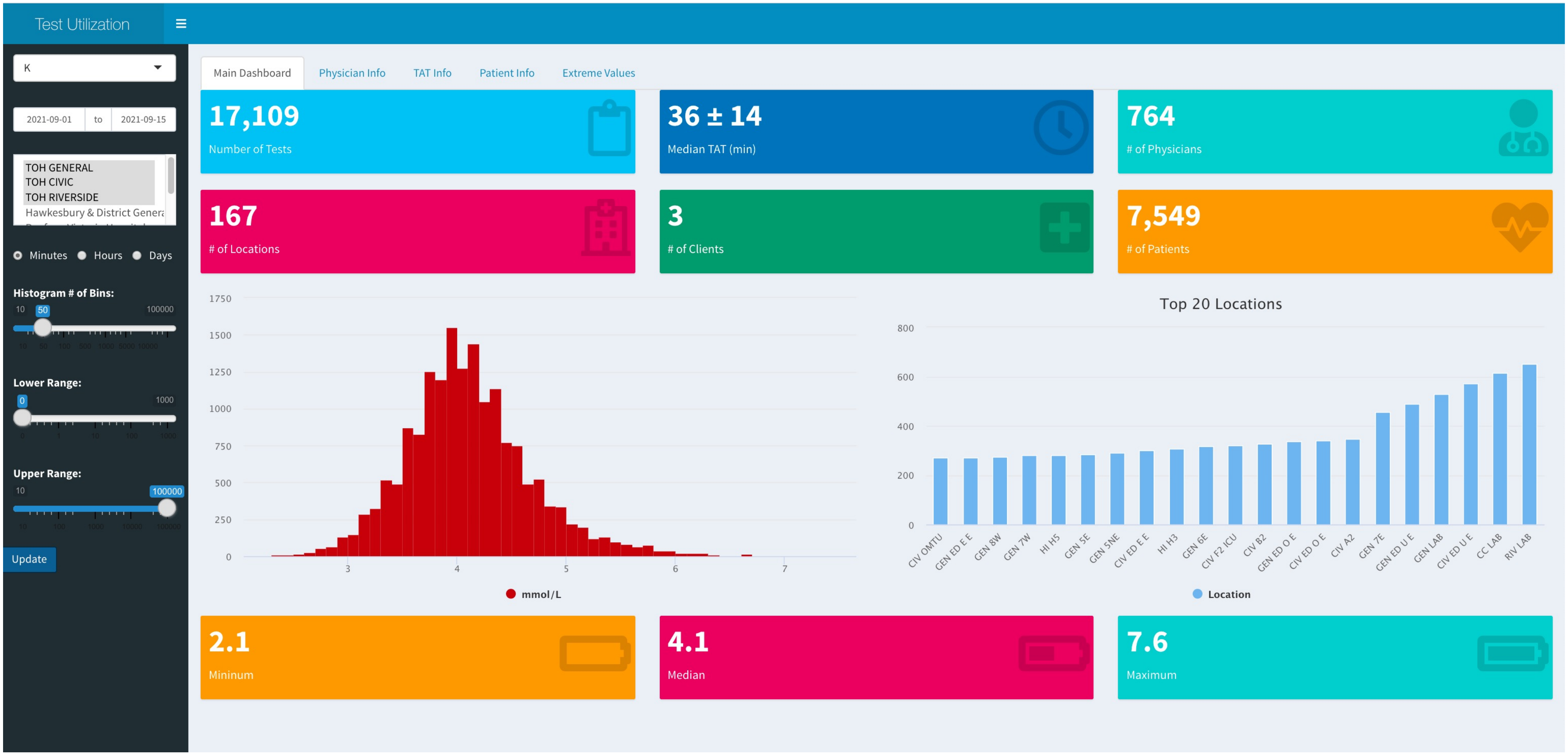
How Many Different Data Sources Do You Use Routinely in the Lab?

A. 0
B. 1
C. 2-3
D. 4-5
E. 6+

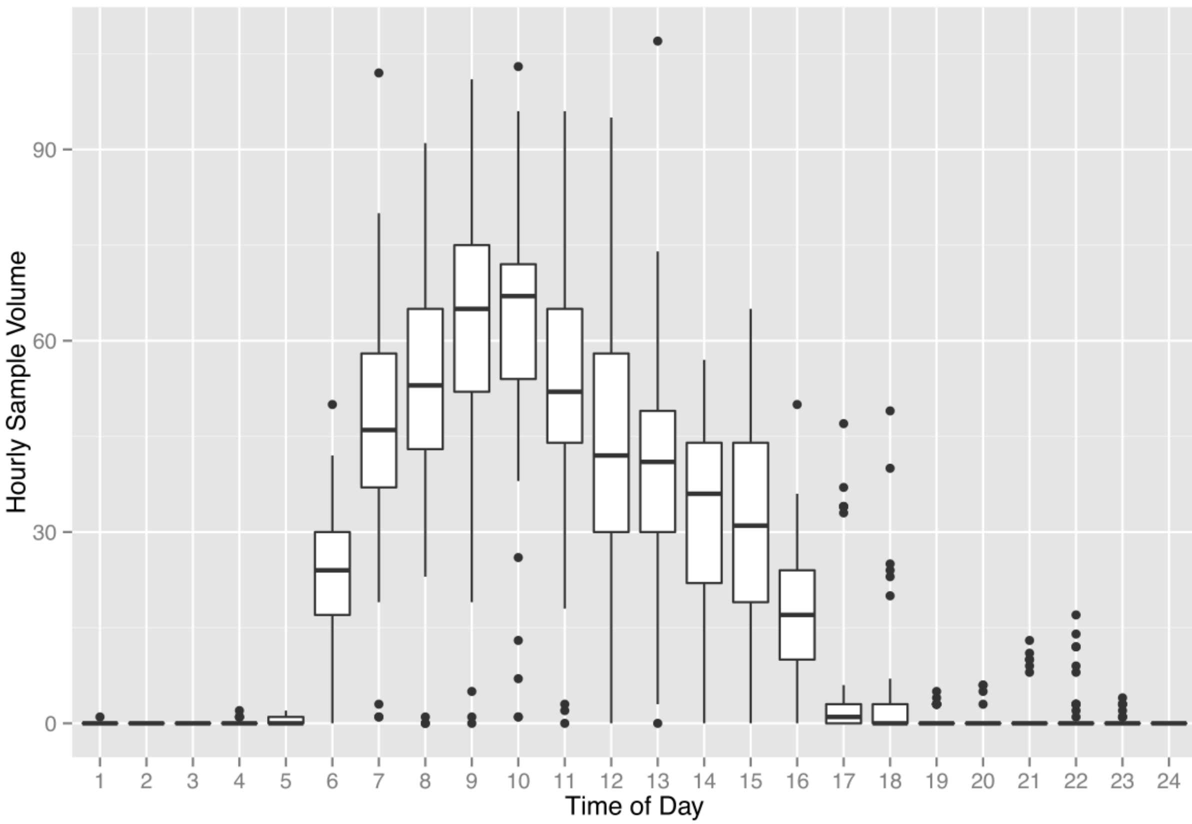
Lab Data Applications

- Understanding of service
  - Ordering, Test volume, Timing, Results, Trends
- Resource Allocation
- Quality
- Generation of New information

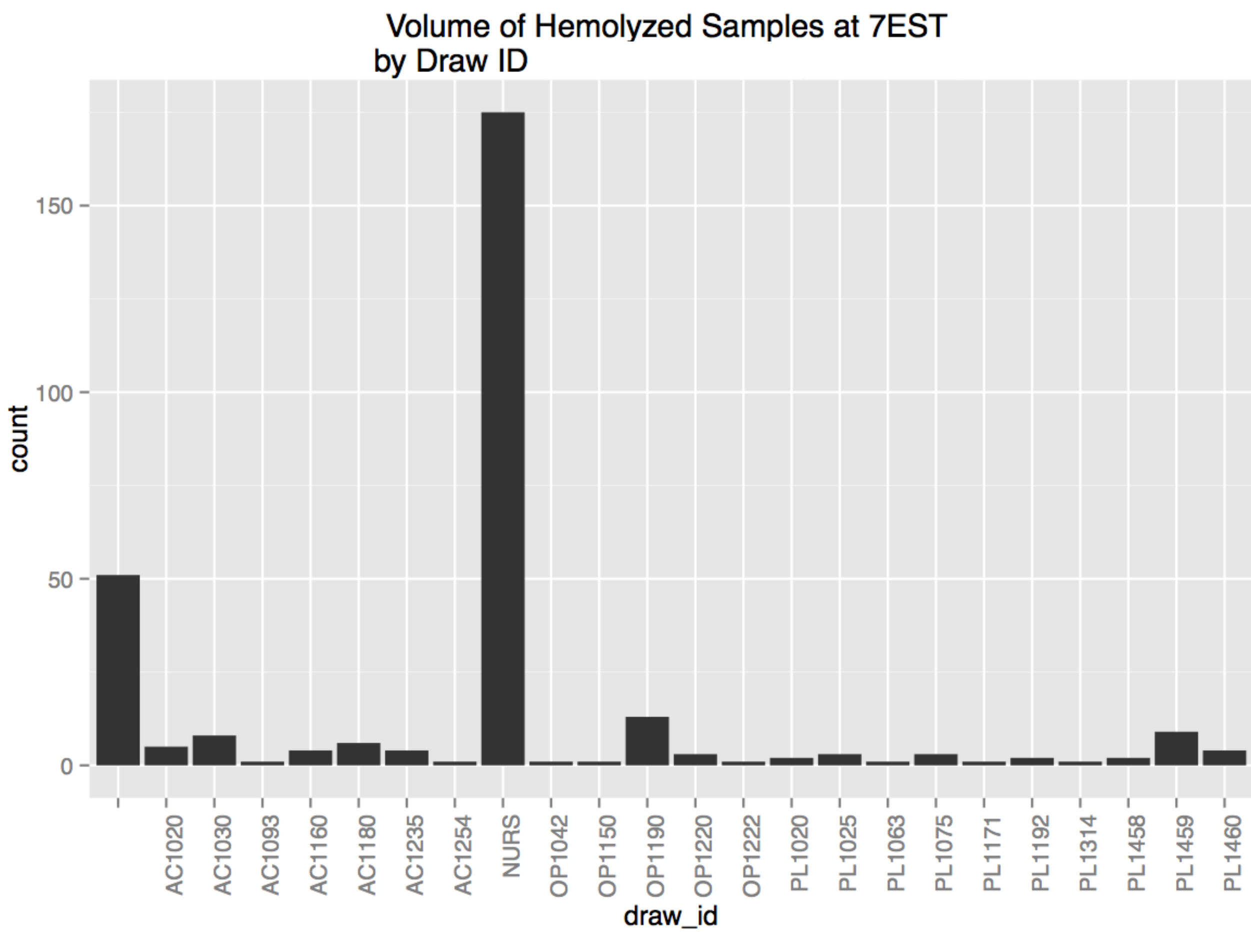
Test Dashboard



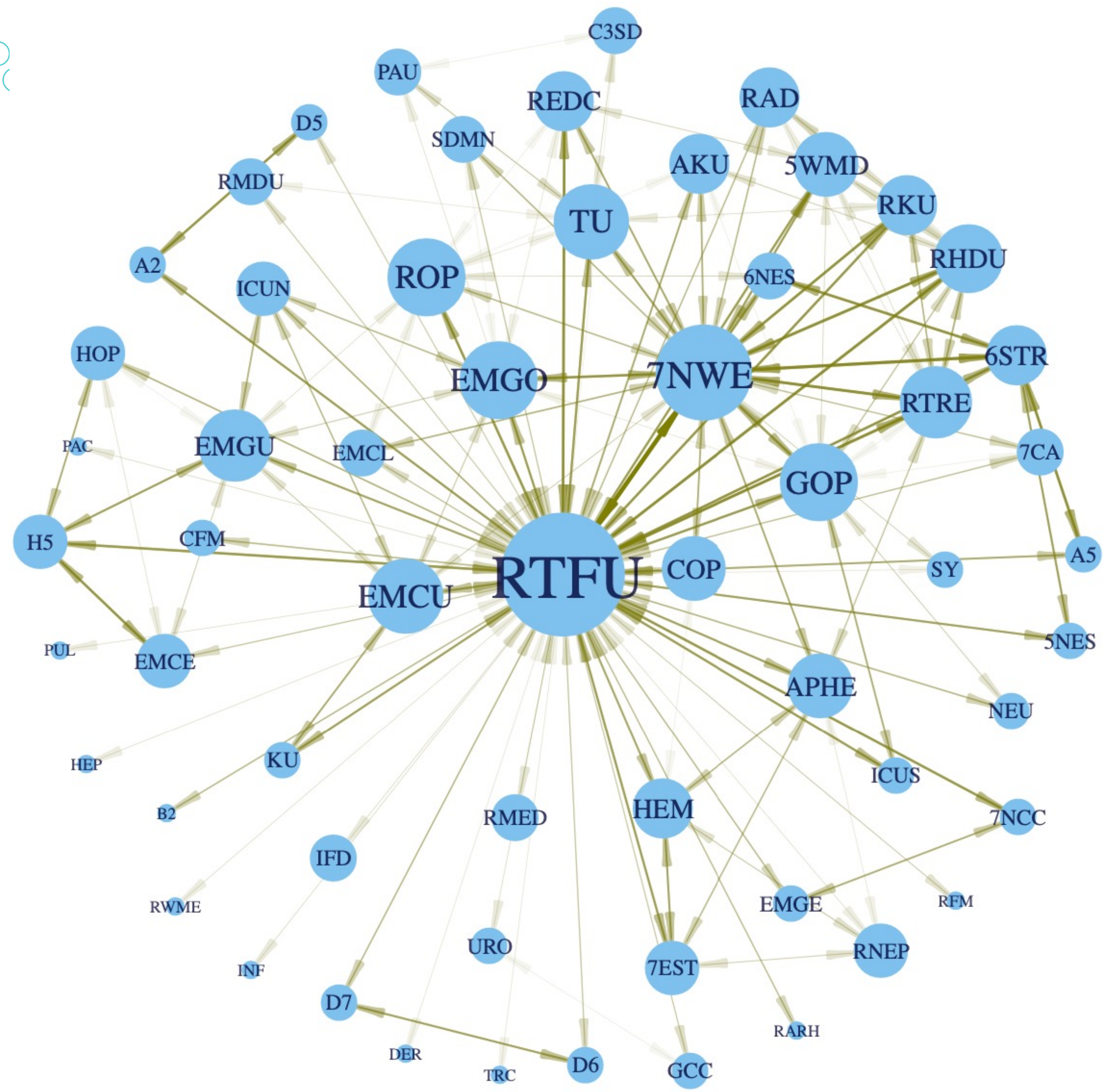
Courier Timing



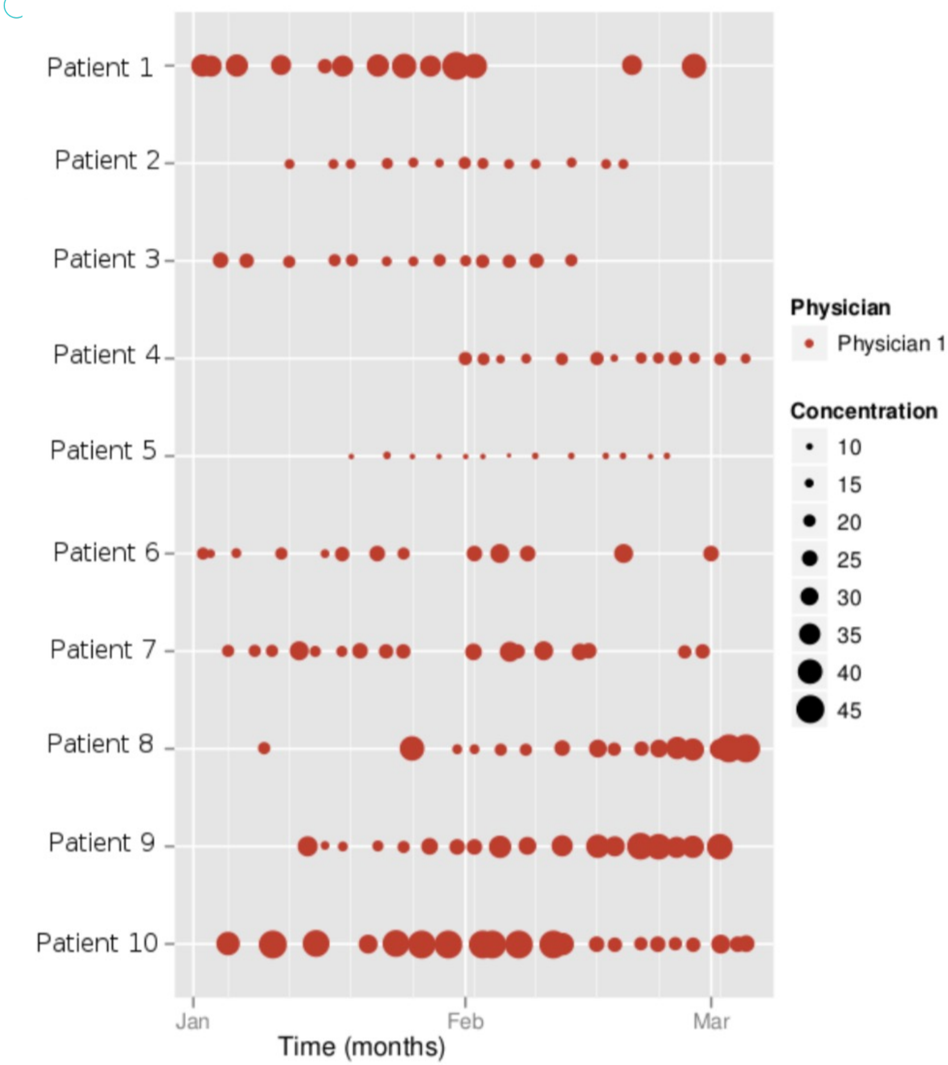
Quality Improvement



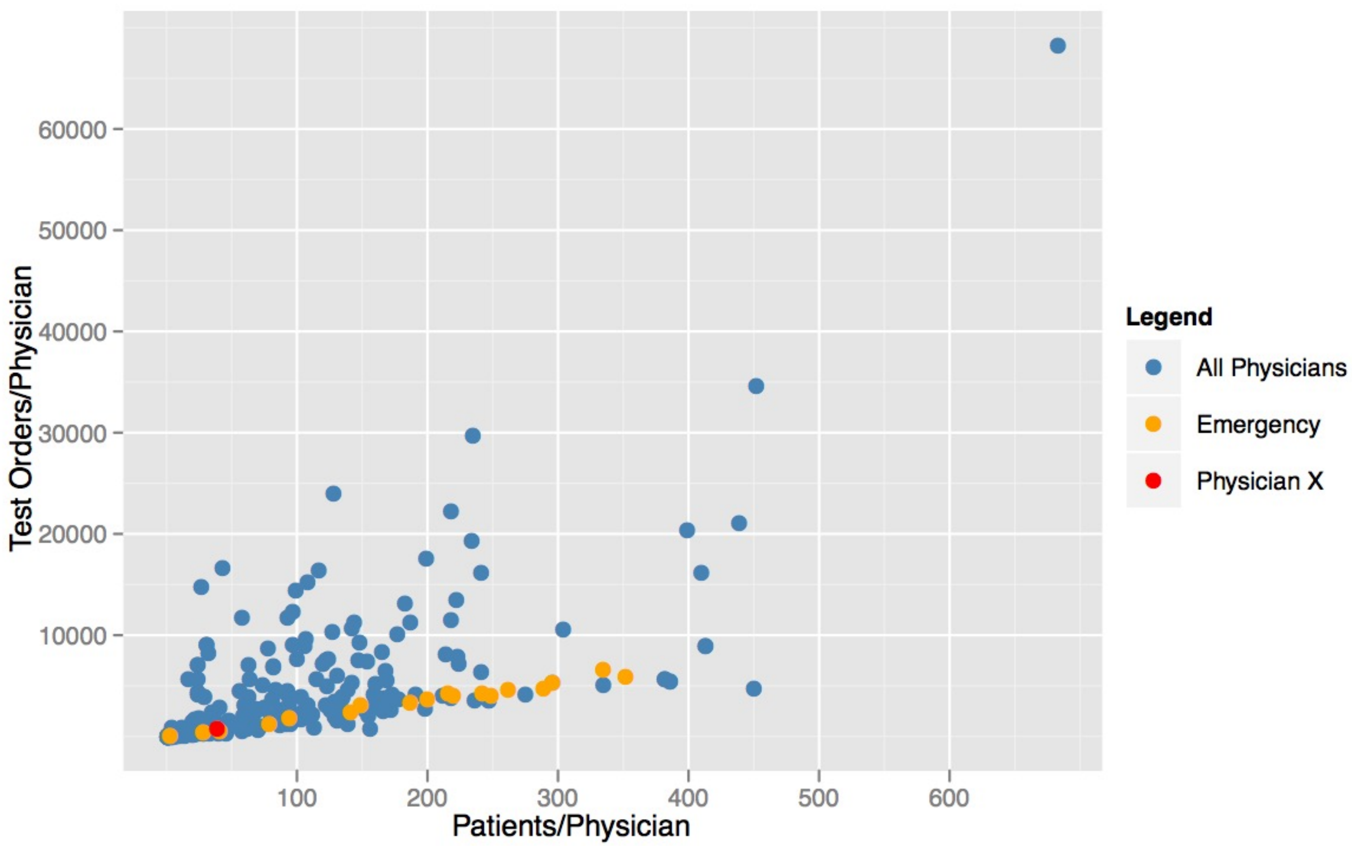
Infection Control



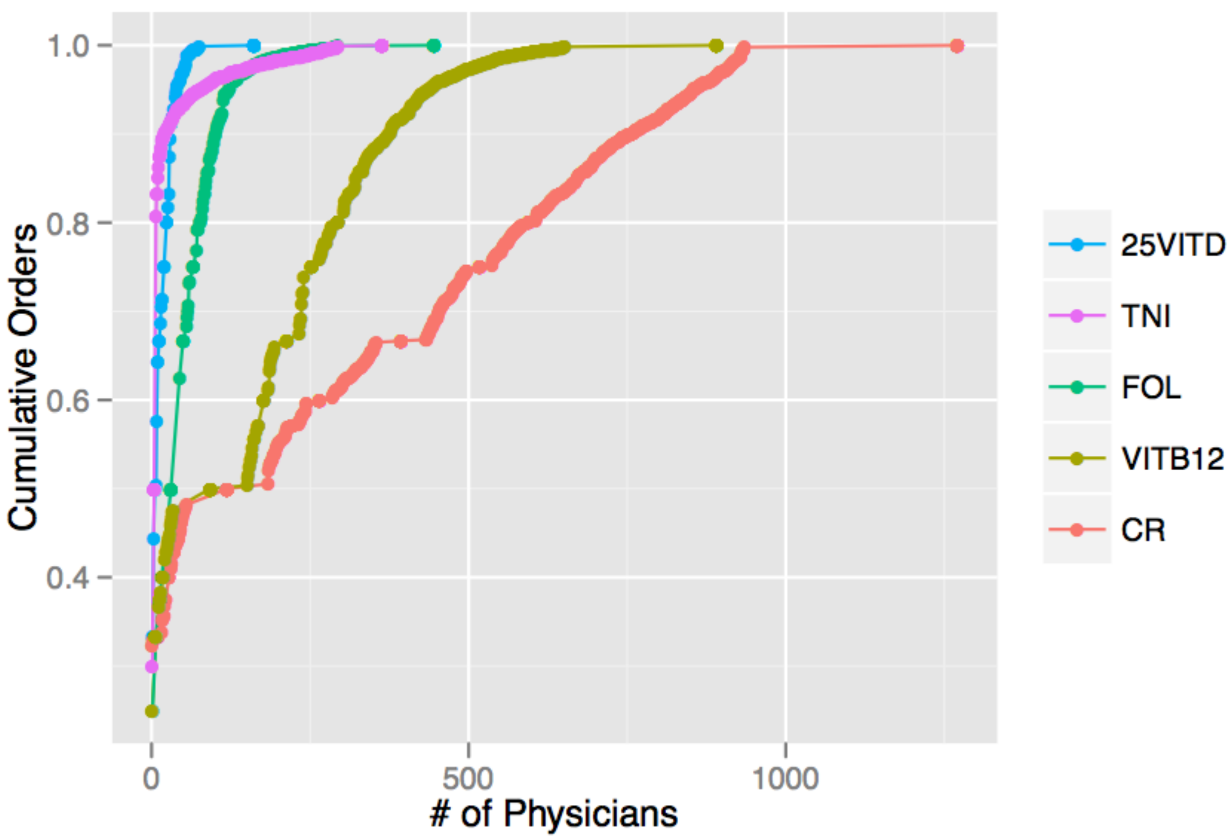
Test utilization



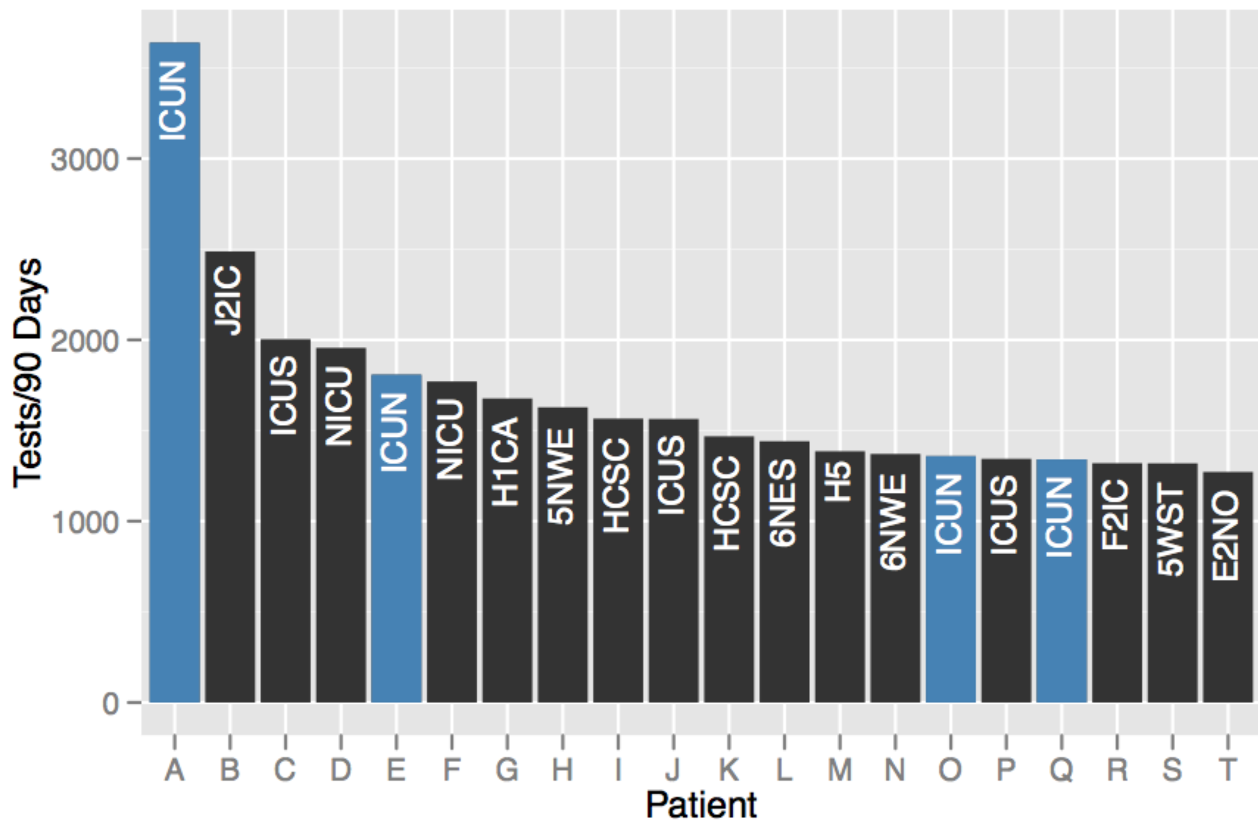
Comparative Test Utilization



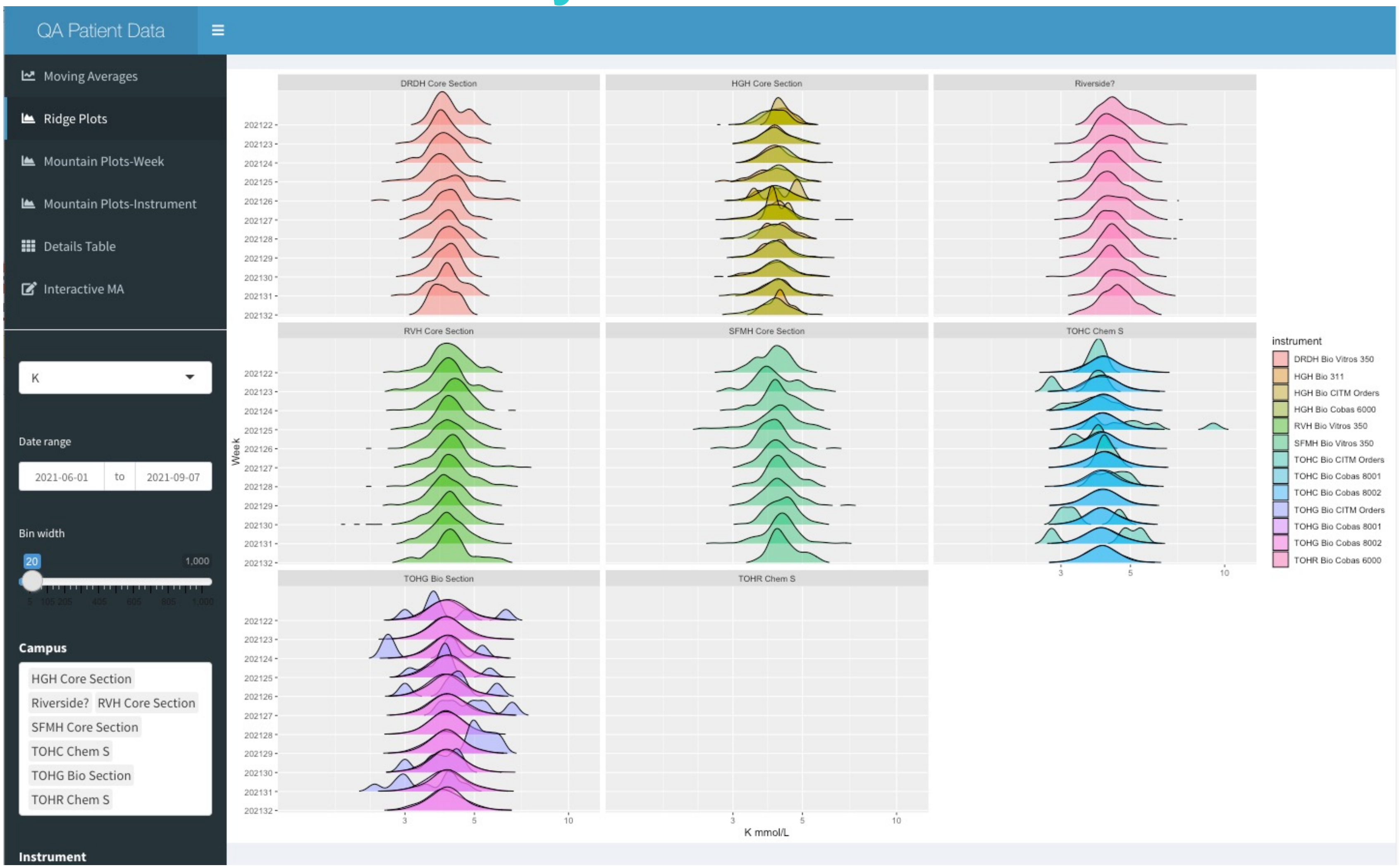
Change Management



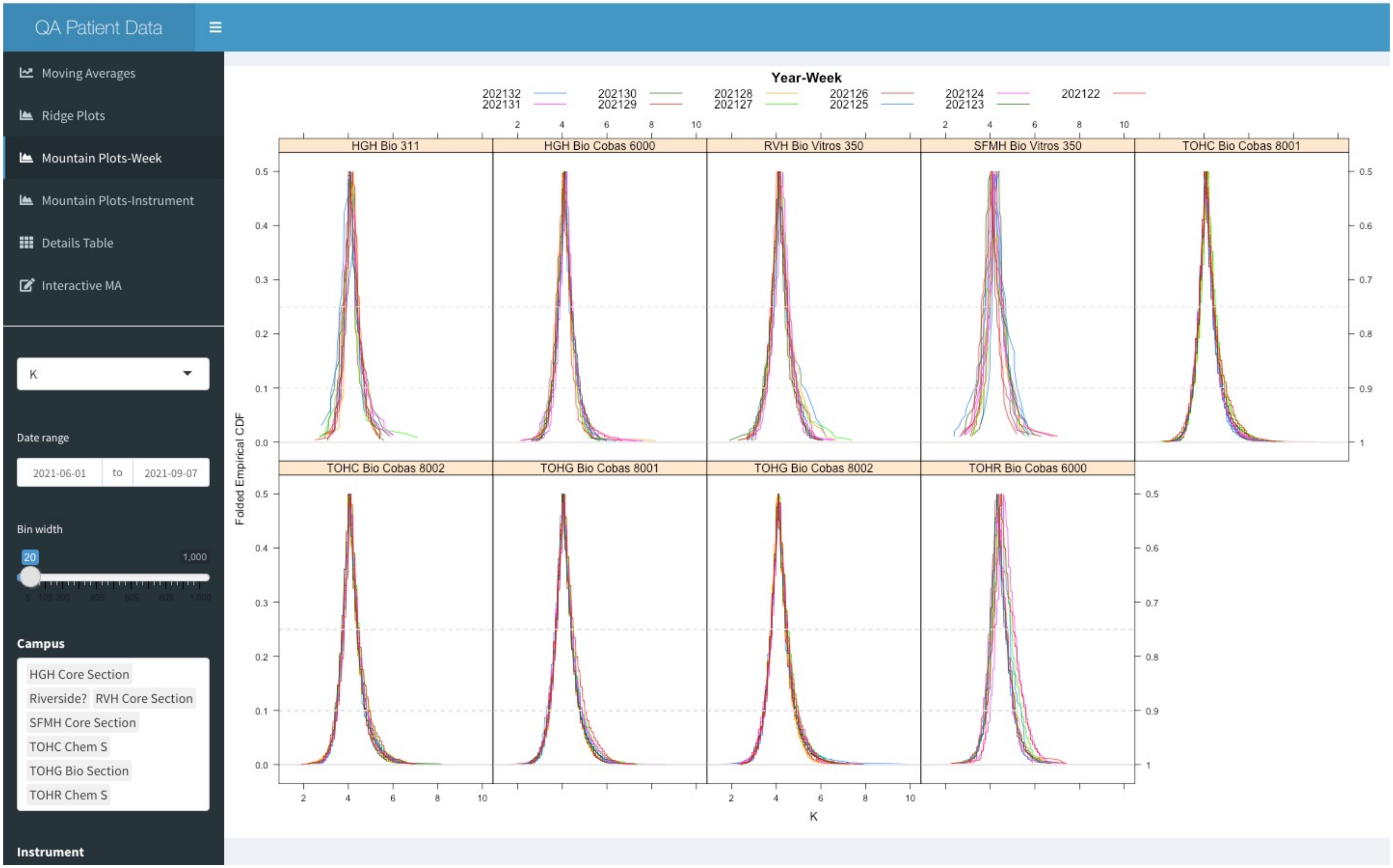
Resource Allocation



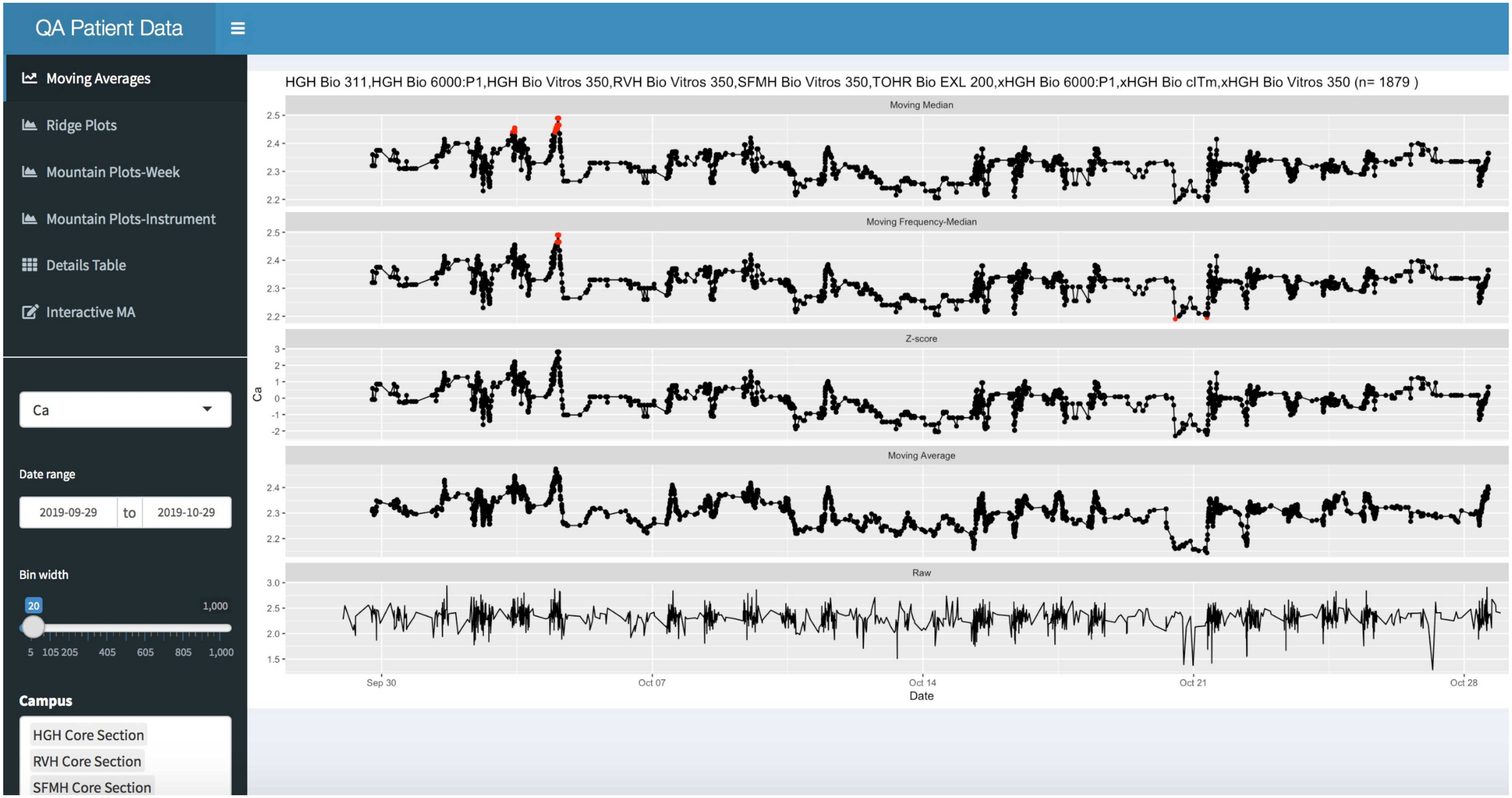
Quality Assurance



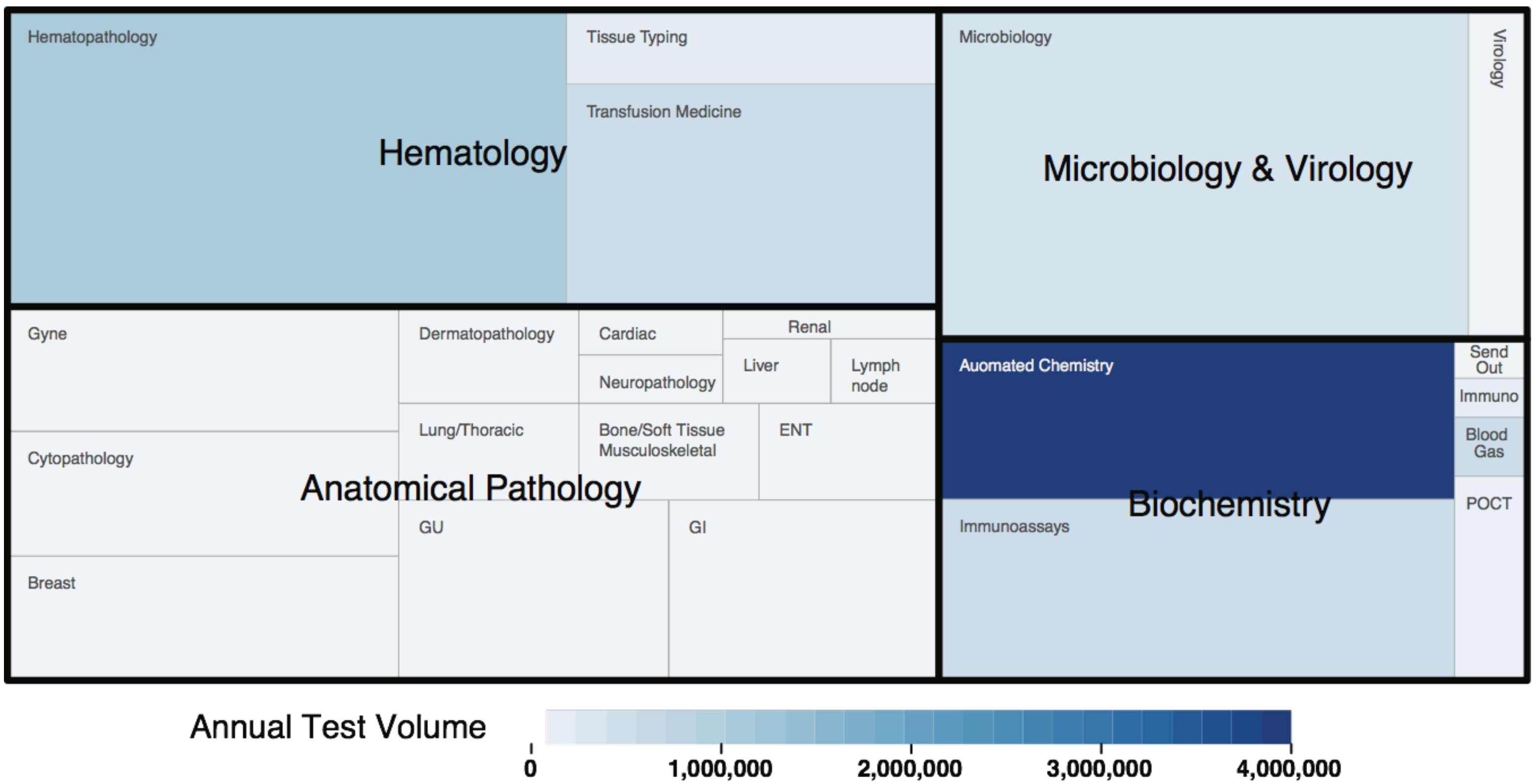
Quality Assurance



Quality Control

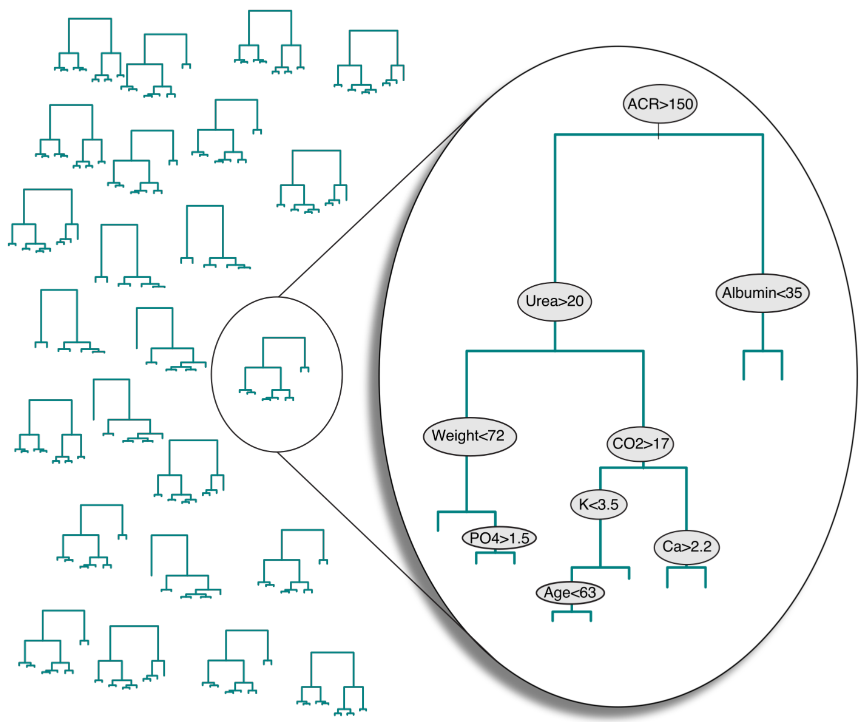


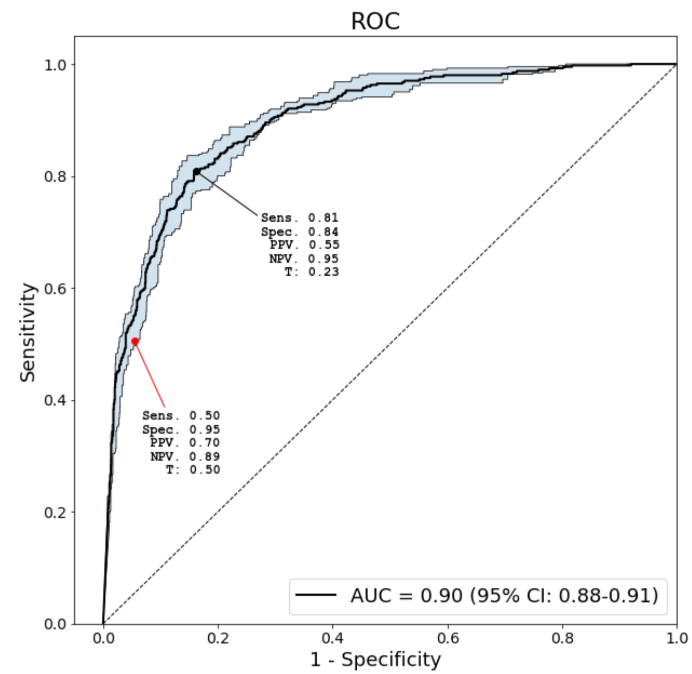
Resource Usage



Cost is proportional to Area

Machine Learning





Part II: Lab Data Challenges

POLL #2

What is the hardest part of using lab data to solve a problem?

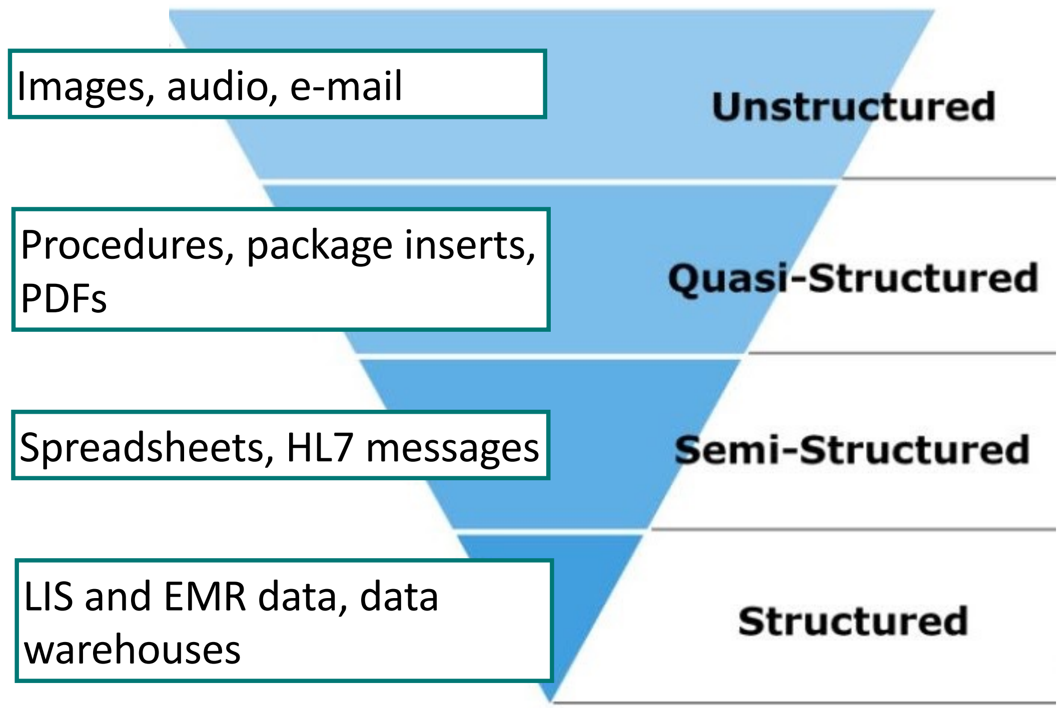
A. Accessing the data
B. Handling the amount of data
C. Data accuracy
D. Analysis
E. Other/Don't know

Lab Data Challenges

- Access
- Standards
- Human resources/Skills
- Interoperability
- Volume
- Validation
- Accuracy
- Software/Hardware Changes
- Ethics/Privacy/Security

- 4 pm, Thursday, September 16, 2021
- 8:00 UTC 16/09/2021
- 2021-09-16 16:00:01
- 2021-09-16 4:00:01 PM
- 1631642896
- 44456.6666782407
- Thurs, Sept 16
- 09-16-21
- 16/09/21
- 16, Sept
- 21-Sep
- 16/09/2021

Lab Data Types



Data Sizes & Approaches

Size	Storage	Software	Scale	Sample
Tiny	Kilobytes	Excel, R	Human readable	Single test method validation
Small	Megabytes	R, SAS, SPSS, Tableau, Python	Fits in computer memory	One week of LIS data
Medium	Gigabytes	Relational database	Fits on a computer	One year of clinical data
“Big”	Terabytes+	Distributed database	Multiple computers	Enterprise health data

Data Challenges Summary

- No once-size-fits-all approach to data

- Many considerations:
  - size, structure, security, standards
  - access, accuracy, analysis

- Ability to manipulate data is essential
  - Right tools for the job
  - A way of thinking

Part III: AI and the Lab

POLL #3

My Knowledge of Machine Learning Is:

A. Non-existent
B. Limited
C. Moderate
D. Extensive
E. My algorithm is answering this for me

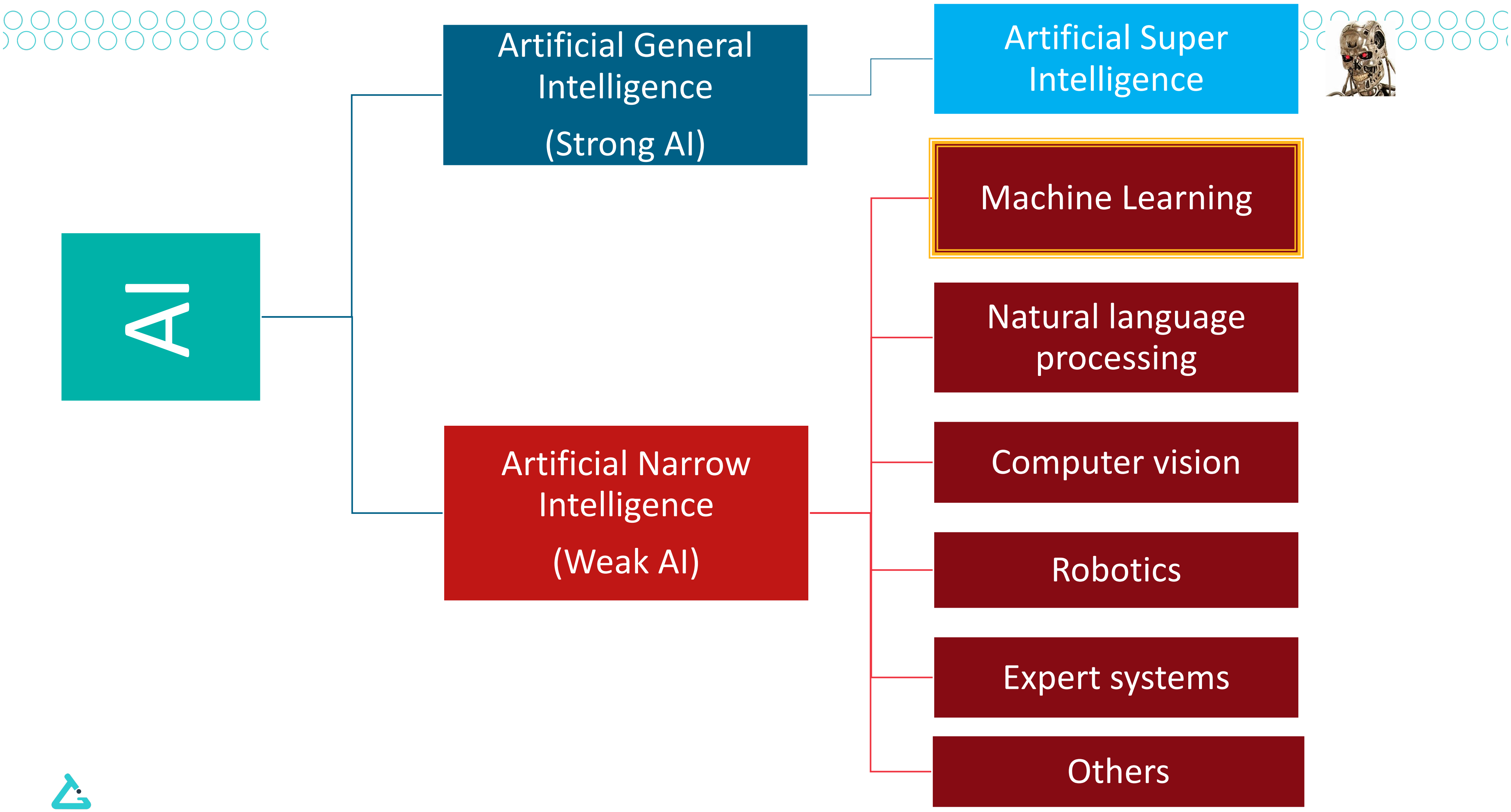
AI

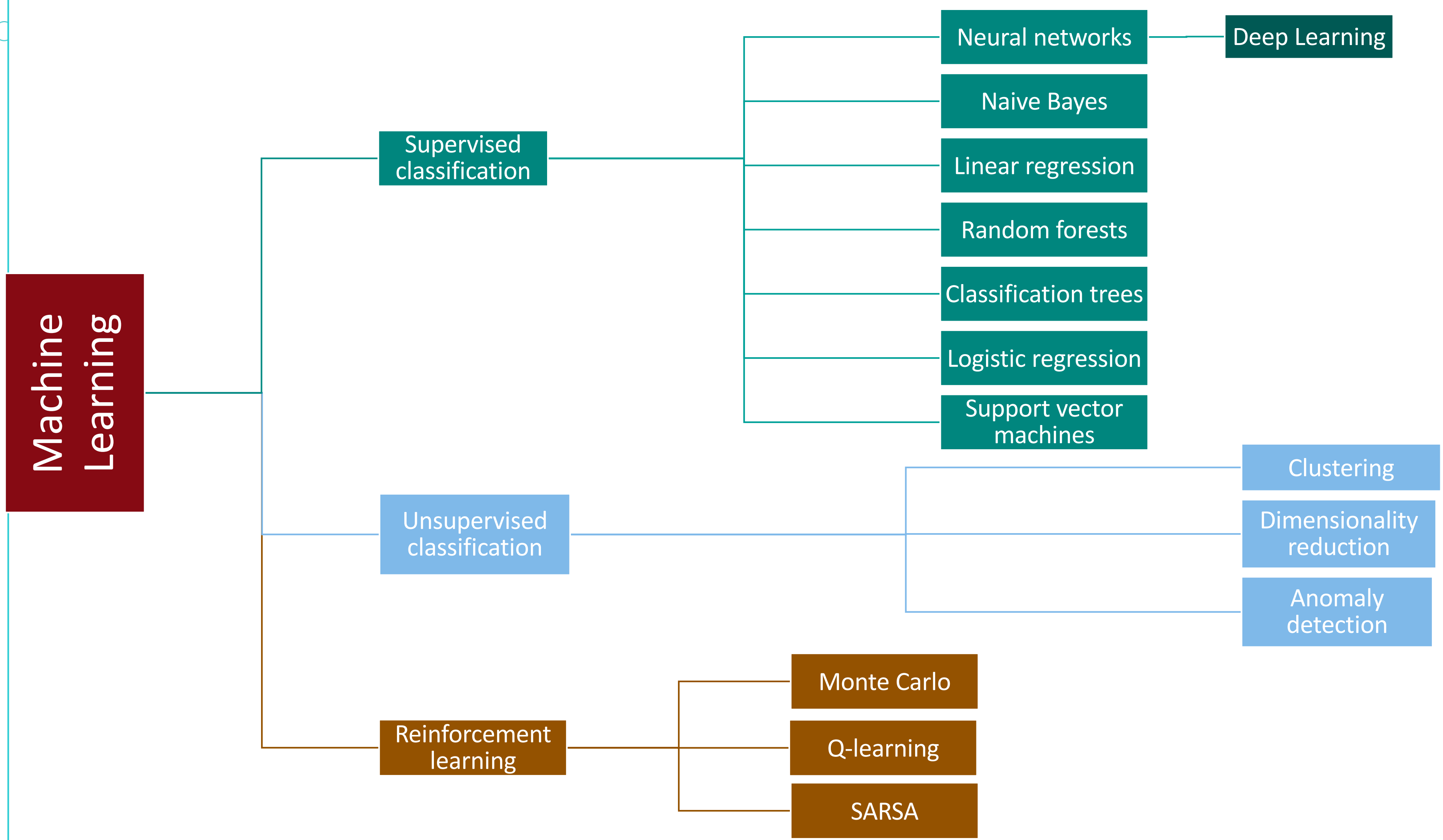
Artificial General Intelligence

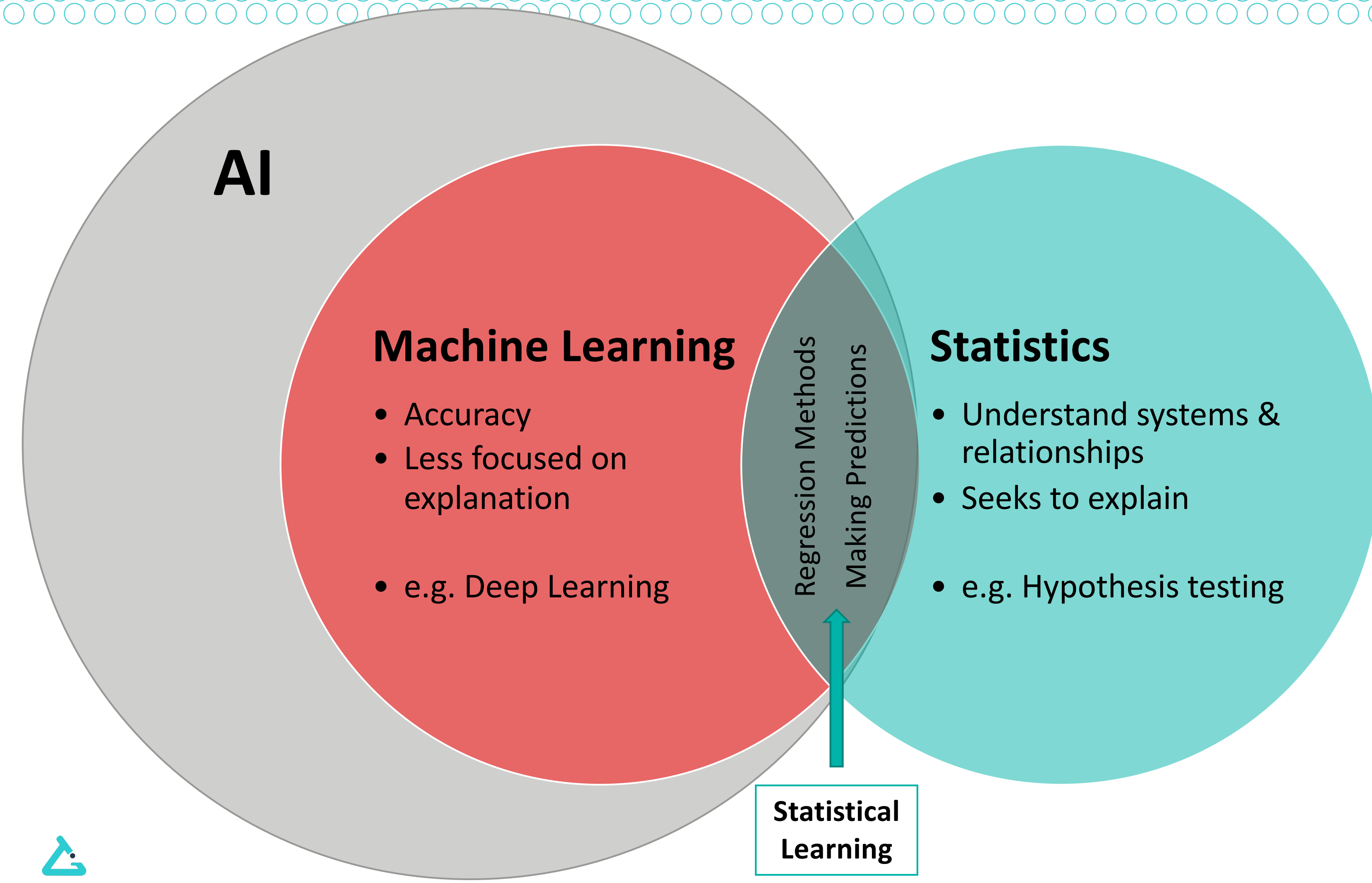
(Strong AI)

Artificial Narrow Intelligence

(Weak AI)


AI
Artificial General Intelligence
(Strong AI)
Artificial Super Intelligence
Artificial Narrow Intelligence
(Weak AI)
Machine Learning
Natural language processing
Computer vision
Robotics
Expert systems
Others


Machine Learning
Supervised classification
Neural networks
Deep Learning
Naive Bayes
Linear regression
Random forests
Classification trees
Logistic regression
Support vector machines
Unsupervised classification
Clustering
Dimensionality reduction
Anomaly detection
Reinforcement learning
Monte Carlo
Q-learning
SARSA


AI
Machine Learning
• Accuracy
• Less focused on explanation
• e.g. Deep Learning
Regression Methods
Making Predictions
Statistics
• Understand systems & relationships
• Seeks to explain
• e.g. Hypothesis testing
Statistical Learning

POLL #4

I have machine learning in my lab

A. Yes
B. No
C. Don't know
D. Don't work in a lab
E. Does a fax machine count?

Machine Learning in the Lab

Growth Areas

- Digital image analysis
  - AP
  - Immunology
  - Hematology
  - UA
- Quality assurance
- Chromatography/MS
- Multi-analyte predictions

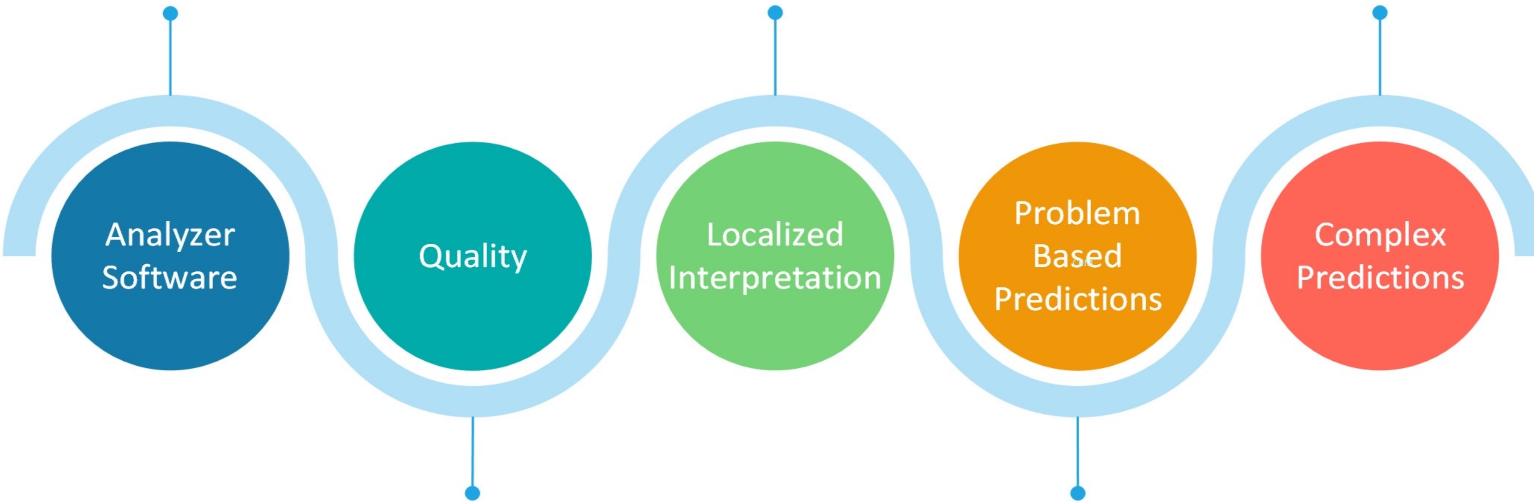
Unlikely near term

- Working with people
- Answering complicated (or poorly articulated) questions
- Teaching
- Strategic planning
- Mission, vision, values
- Performance evaluations
- Regulatory

- Digital image analysis
- Integrated interpretations

- Stand alone applications
- Chromatography & MS
- Electrophoresis
- Panel interpretations

- Patient Wellness
- Early Disease Identification
- Medication recommendations



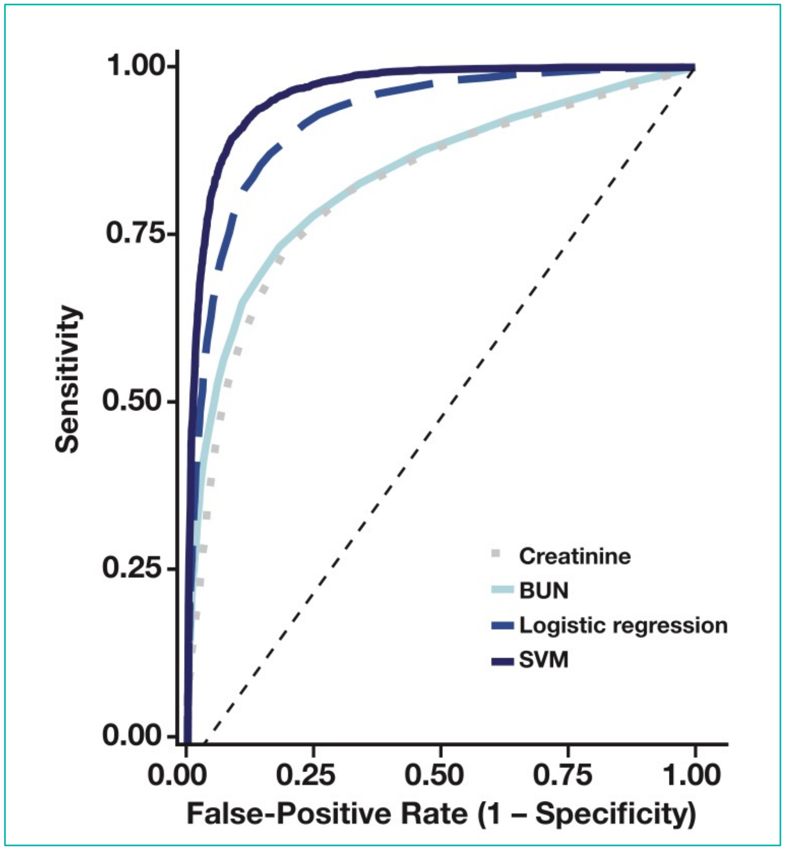
- Error detection
- Real time monitoring
- Stewardship
- Reference intervals

- Disease predictions:
- Chest pain
- COVID
- Sepsis
- AKI
- Acute care

Test Reporting

Neural Network for autoverification

Machine Learning for wrong tube



Quality Review of MS Data

Rosenbaum & Baron Am J Clin Pathol 2018;150:555-566

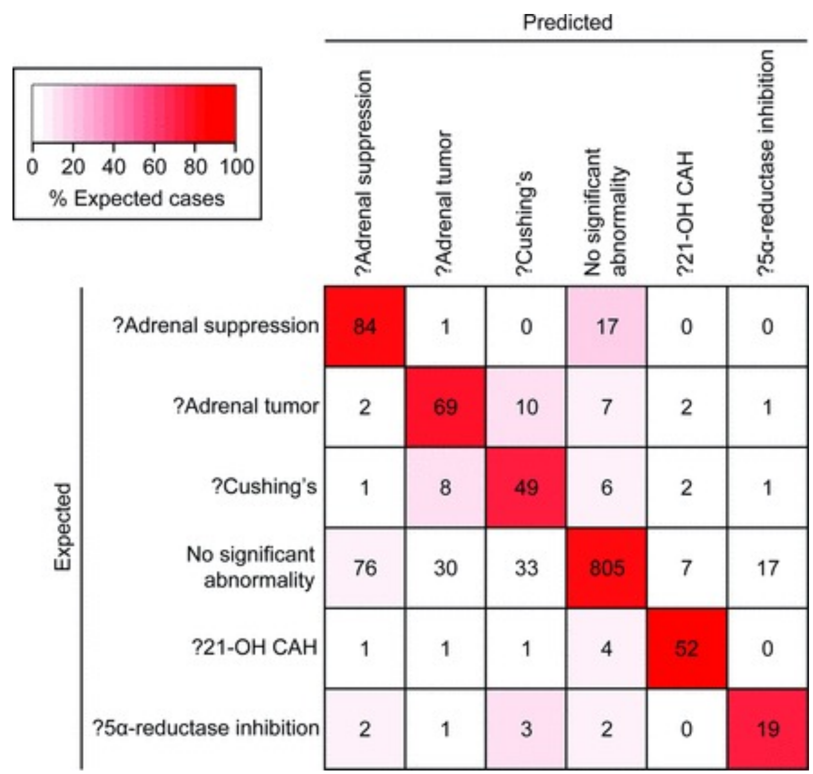
Result Interpretation

Plasma amino acid interpretation

Urine steroid profiles

Acute kidney injury prediction

COVID-19 prediction



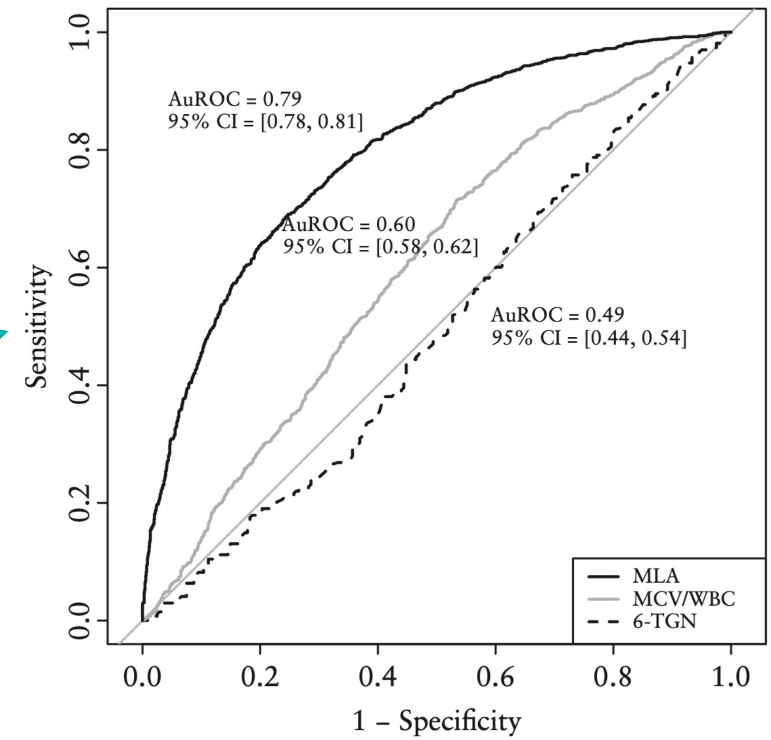
Wilkes et al. *Clinical* Chemistry 64:11, 2018

Machine Learning as a Diagnostic Test

Thiomon:

- Prediction of thiopurine treatment success in IBD Patients
- CBC and common chemistry analytes to predict:
  - Response to thiopurines
  - Shunting
  - Compliance

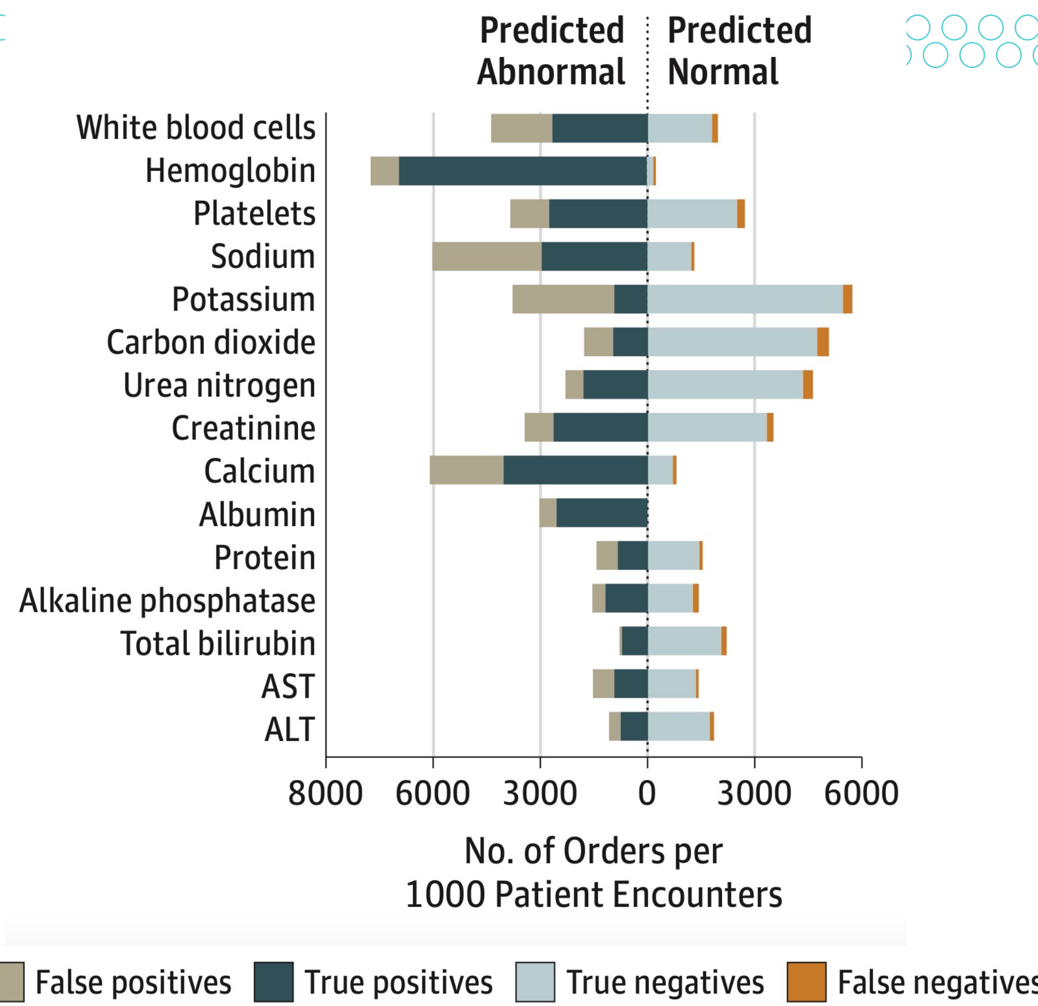
Predicting intestinal inflammation remission



Waljee et al. Journal of Crohn's and Colitis, 2017, 801–810

Test Utilization

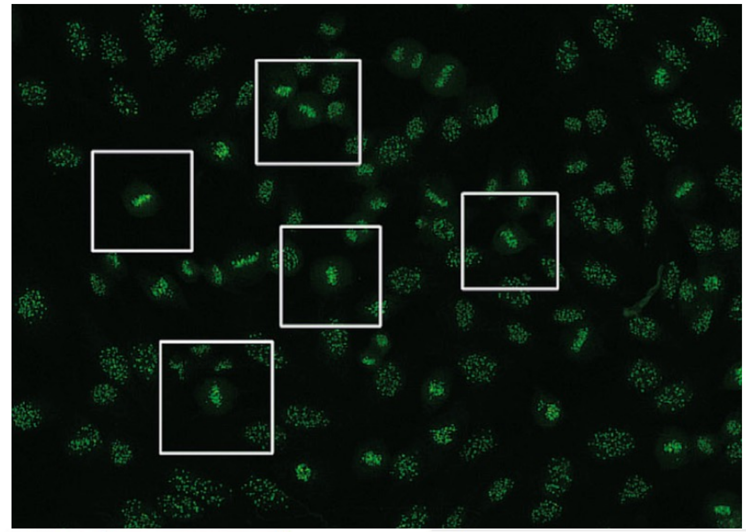
- Prediction of Low-Yield Tests
  - Inpatients
- LFTs
- Iron deficiency

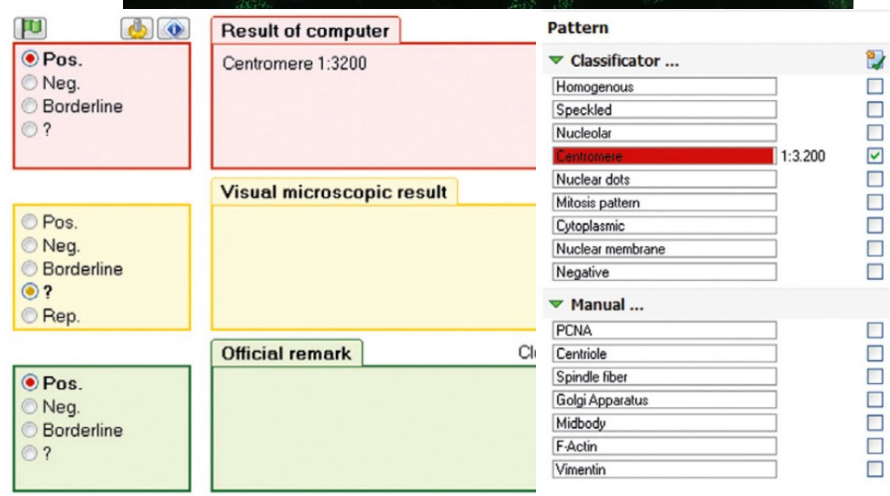


Xu et al. JAMA Network Open. 2019;2(9):e1910967.

Automated digital image analysis

- Computer-aided immunofluorescence microscopy
- Automated Urine microscopy
- Peripheral blood smear review





Final Note: Machine Learning in Healthcare Cautions

- Data, Design, & Bias
- Ethics & Legality
- Privacy & Security

Summary

- Lab Data is abundant and rich
- Harvesting and cleaning data can be difficult
- Machine learning is another tool for problem solving
- Treats algorithms as unproven tests

References and Suggested Readings

- Rosenbaum & Baron Am J Clin Pathol 2018;150:555-566
- Waljee et al. Journal of Crohn's and Colitis, 2017, 801–810
- Clinical Chemistry 66:9 1210–1218 (2020)
- Circulation. 2019;140:899–909
- https://arxiv.org/pdf/1803.03453v1.pdf
- Biol Infect. 2020 Oct;26(10):1310-1317.
- BMC Medical Informatics & Decision Making V19:171 (2019)
- http://dx.doi.org/10.1016/j.tim.2016.12.006
- Ann Lab Med 2019;39:15-22
- Front Immunol. 2018; 9: 927.
- https://doi.org/10.1177/0961203314559635
- Clinical Chemistry, 63:12, 1790–1791 (2017)
- Clinical Chemistry 65:10 (2019)
- FDA Guidance documents:
  - Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback. FDA.