NAM Digital Learning Collaborative:

Practical Challenges for AI/ML Development, Spread, and Scale

Michael E. Matheny, MD, MS, MPH

Director, Center for Population Health Informatics
Associate Professor, Departments of Biomedical Informatics, Medicine, and Biostatistics
Vanderbilt University Medical Center

Associate Director for Data Analytics, VINCI
Associate Director, Advanced Fellowship in Medical Informatics
Tennessee Valley Healthcare System VA

Twitter: @MichaelEMatheny
Email: michael.Matheny@va.gov, michael.Matheny@Vanderbilt.edu

CENTER FOR POPULATION HEALTH INFORMATICS

Model Explainability

- Machine learning methods
  - Lauded for achieving higher accuracy than traditional statistical methods
  - Criticized for being ‘black boxes’ that are not explainable or interpretable

- One of the key limitations in scale and spread of ML methods has been in areas where transparency and explainability are needed

Rise of Demand for Explainability

- Pivotal Moment: European Union passing General Data Protection Regulations that specifically cover requirements for transparency in automated modeling set to be law in 2018

- DARPA Explainable Artificial Intelligence (2016) RFA/Announcement

Goodman, et al. arXiv:1606.08813
https://www.darpa.mil/program/explainable-artificial-intelligence

Explainable to Whom? How? For What?

- What is explainability, <u>really</u>?
- Whole notion of allowing data science to determine interactions and relationships in the data that are not discernable to humans is core to the discipline
- Not all models have to be interpretable to achieve the desired effect
- Complex ‘traditional statistical’ models are not necessarily easily explainable

Explainability

- Who needs the explanation, and for what:
  - Medical Professional
    - needs to know what is modifiable in the clinical context
    - Needs to know likely impact of using the data - change in outcome(s)
  - Patient
    - needs to trust that the model is not discriminatory (race, SES, gender, sexual orientation, sensitive clinical conditions, etc)
    - Needs to be able to which of their characteristics were important in the modeling and resulting care decisions
  - Medical Societies
    - Needs to know that care supported by modeling is consistent with best practice guidelines and care recommendations

Active Areas of Research

- Methods for visualizing model outputs in a way that are clinically interpretable
- Use of simpler models on top of ML to provide interpretation
- explaining individual prediction versus overall model (parameters, etc)

Overall Model Performance

- How much accuracy is needed for clinical implementation?

- Application & Context Dependent
  - Population Health Care Management
  - Point of Care Management

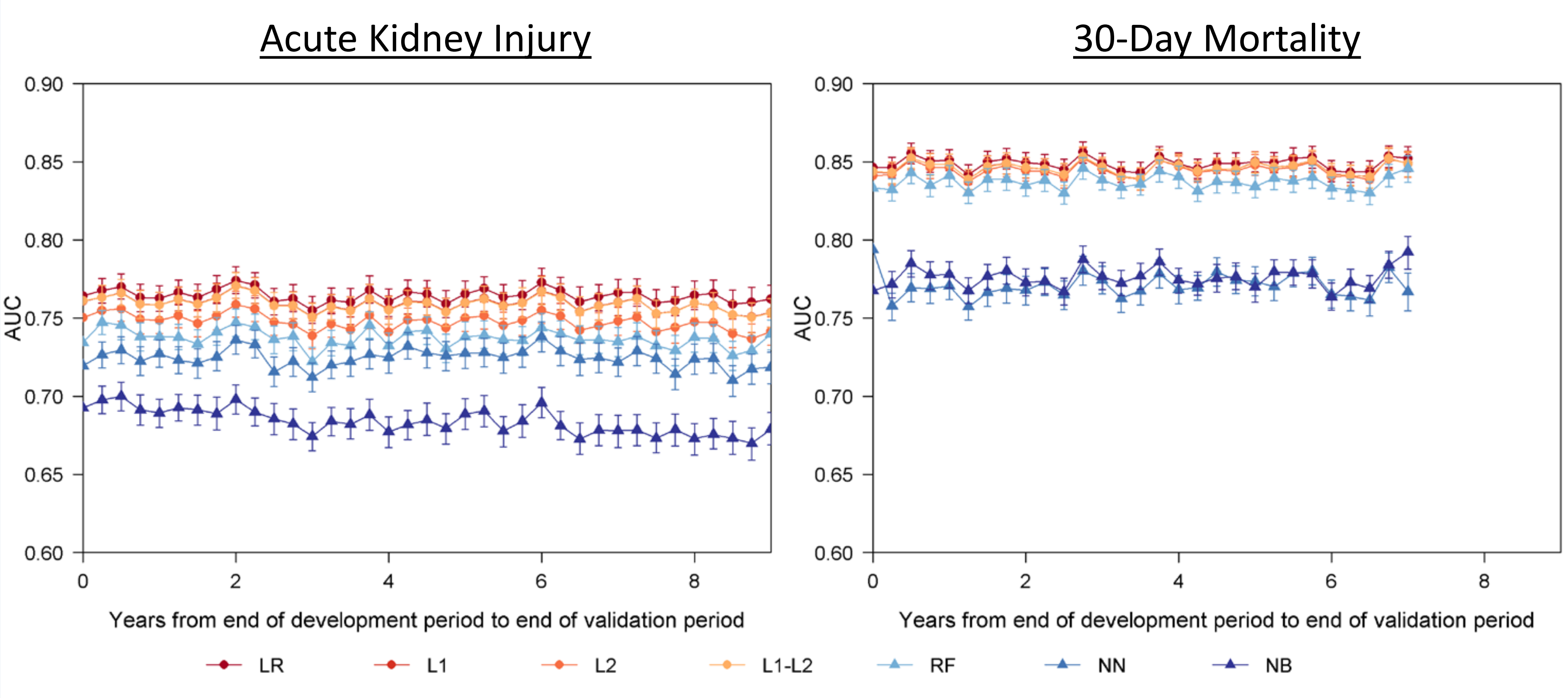
- Knowing when it is sufficient to use a model to separate a ‘yes’ from a ‘no’ (discrimination) and when do you need a range of probabilistic predictions (calibration)

Challenges in Re-Use (Spread) of Models

- Immediately, external application of models face:
  - Variation in underlying data collection
  - Variation in implementation of features
  - EHR Implementation: Real-Time Data changes ('dirty production data')

- Performance Over Time:
  - Changes in clinical practice
  - Changes patient case mix and outcome rates
  - Changes in data collection practices

Discrimination Over Time



Davis SE, et al, Matheny ME. J. Am. Med. Inform. Assoc. 2017 Nov;24:1052-1061.

Davis SE, et al, Matheny ME. AMIA Annu Symp 2017

Calibration Over Time: Estimated Calibration Index