# The model-to-data paradigm: overcoming data access barriers in biomedical competitions

Thomas Schaffter<sup>1,\*</sup>, Timothy Bergquist<sup>2,\*</sup>, Yao Yan<sup>3</sup>, Thomas Yu<sup>1</sup>, Vikas Pejaver<sup>2</sup>, Noah Hammarlund<sup>2</sup>, Justin Prosser<sup>4</sup>, Sean Mooney<sup>2</sup>, Justin Guinney<sup>1,2</sup>

Model-to-data paradigm enables researchers to train and evaluate models on critical data derived from health care and clinical trials

**Alternative models for sharing confidential biomedical data** 

Critical patient information derived from academic research, health care and clinical trials are off limit for traditional data-to-model challenges. Existing barriers include:

Academic

Research

• Access to big and sensitive data



Health Care

+

**Clinical Trials** 

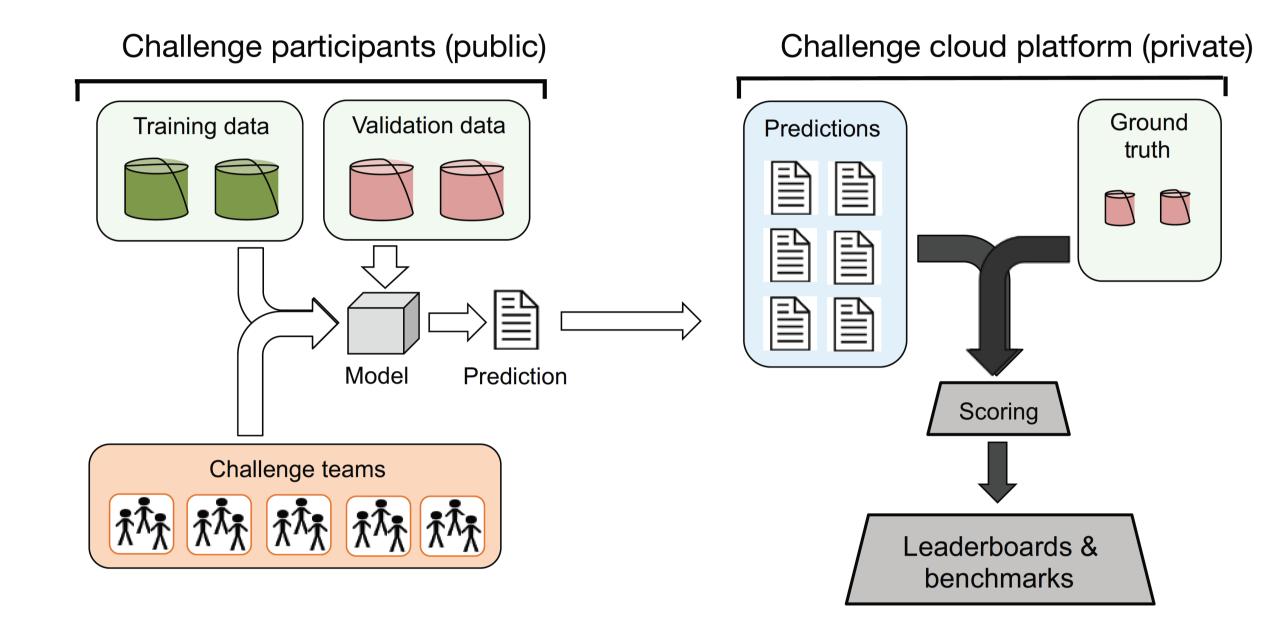
**EHR DREAM Challenge: Patient Mortality Prediction** 

Our vision is to apply the MTD paradigm to build a network of EHR Data Nodes to enable participants to develop predictive models.

For this first EHR Challenge, participants are addressing the following question:

Lack of effective frameworks for assessing performance & generalizability

# **Traditional data-to-model challenges**



#### **Our solution: The model-to-data paradigm**

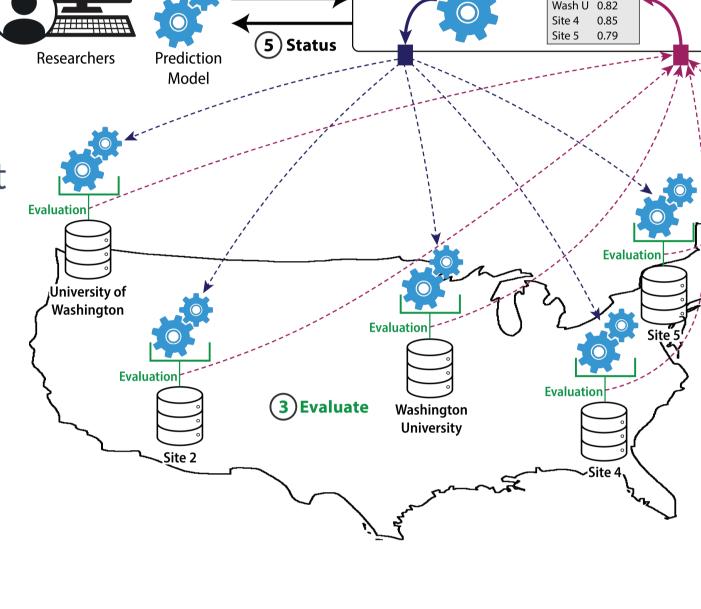
Containerized models submitted by participants are trained and/or evaluated on sensitive data in secure environments. The reproducibility of the results is also easier to achieve.

Given all the past EH Records of a patient, predict the probability that the patient will pass away within 180 days following his/her most recent exam.

# **Challenge data**

- Synthetic data (SynPUFF)
- University of WA EHR data warehouse
  - 1.3 million patients with at least one visit
  - 22 million visits
    - 5 million drug exposure
    - 10 million observations
    - 48 million conditions
    - 221 million measurements
  - Format: OMOP common data model

# Submission workflow

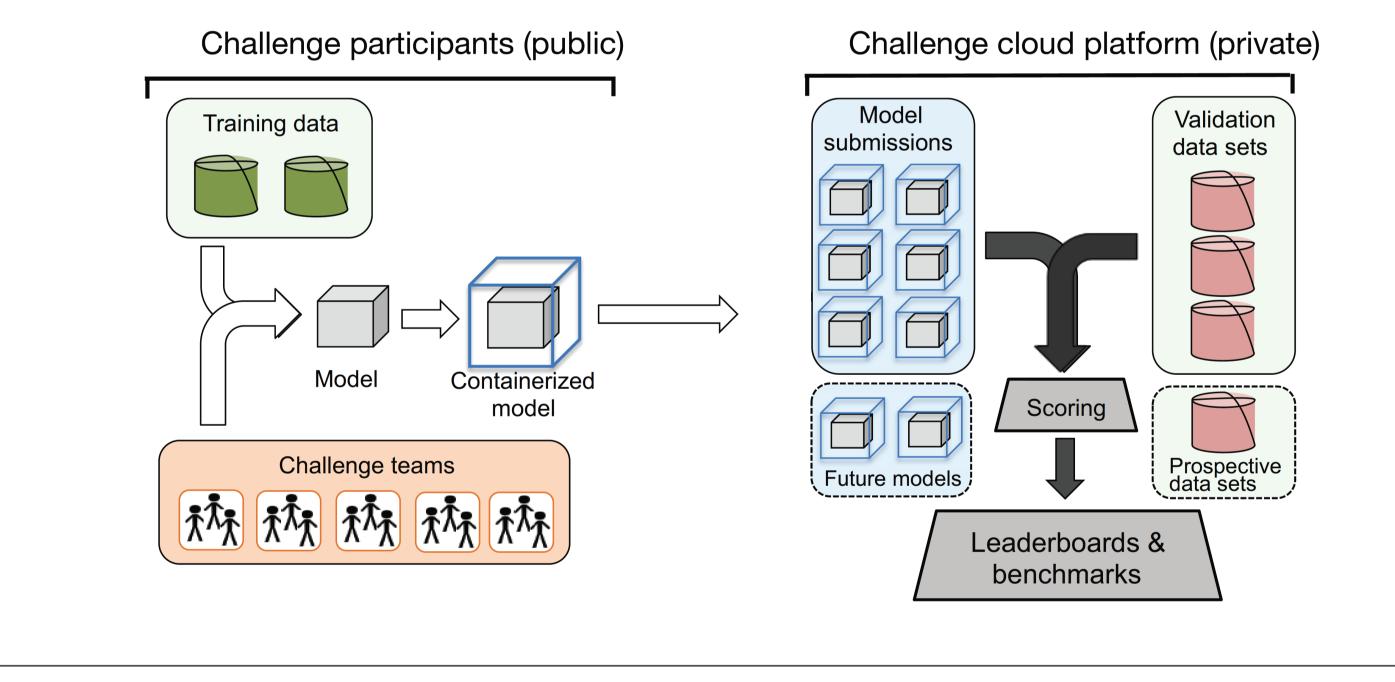


(1)Build

Central Model Repository

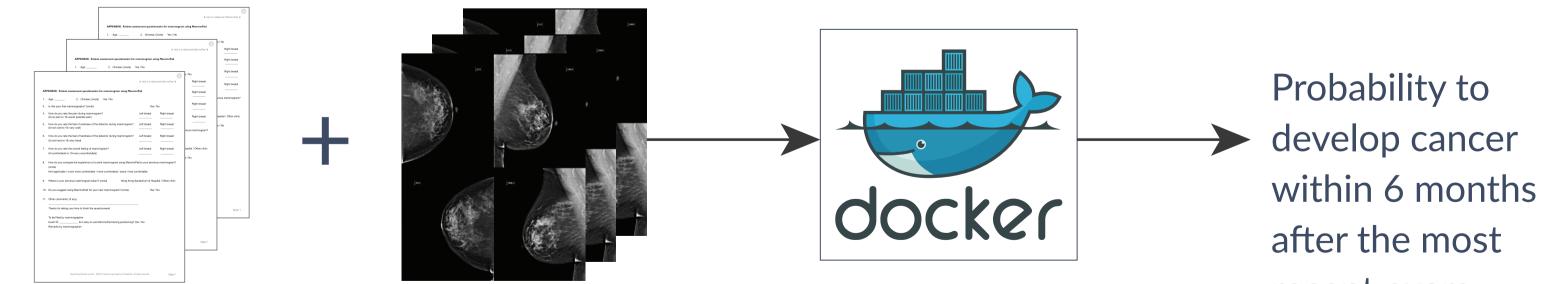
0.92

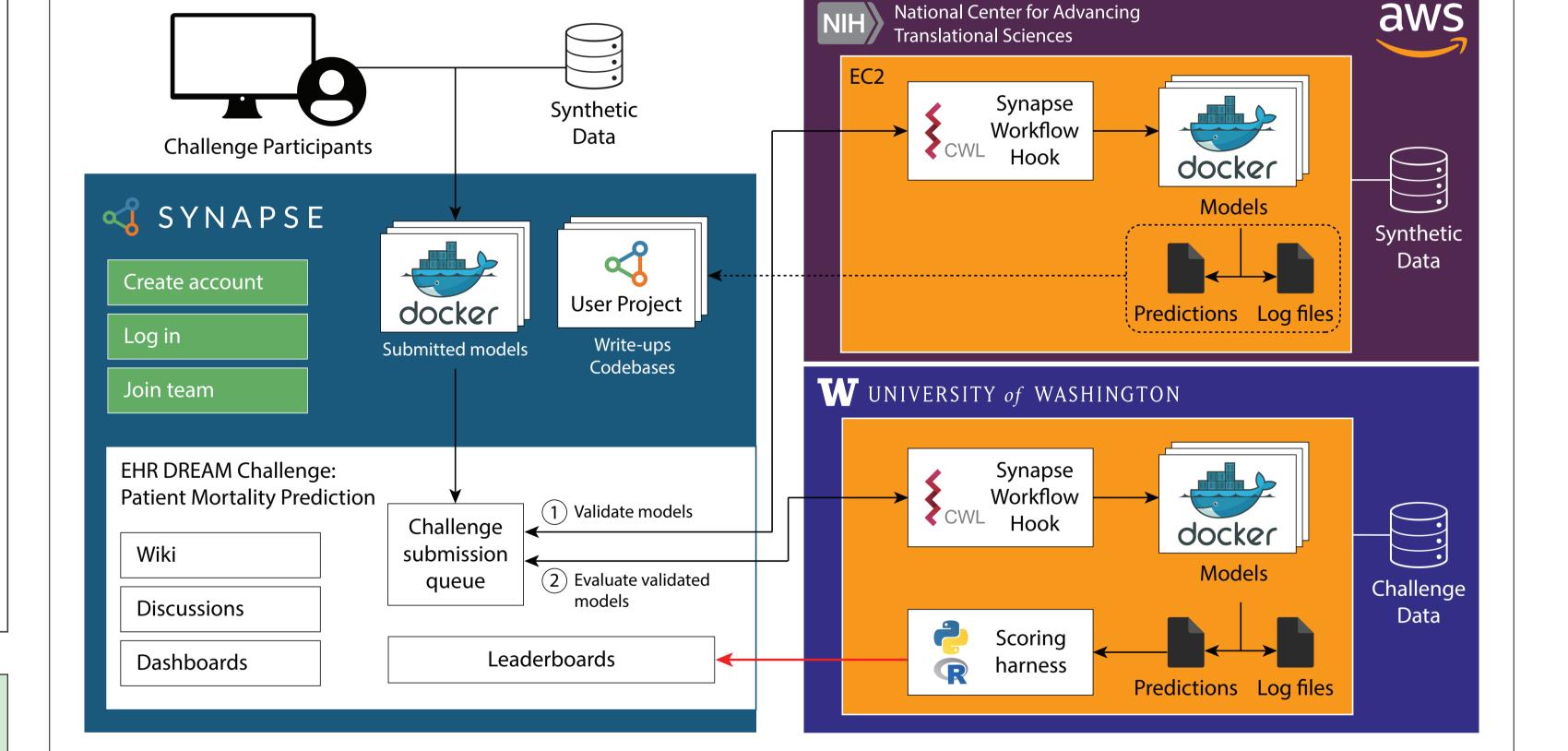
Score



# **Model-To-Data DREAM Challenges**

The MTD approach has already been implemented in several DREAM Challenges such as The Digital Mammograph Challenge.

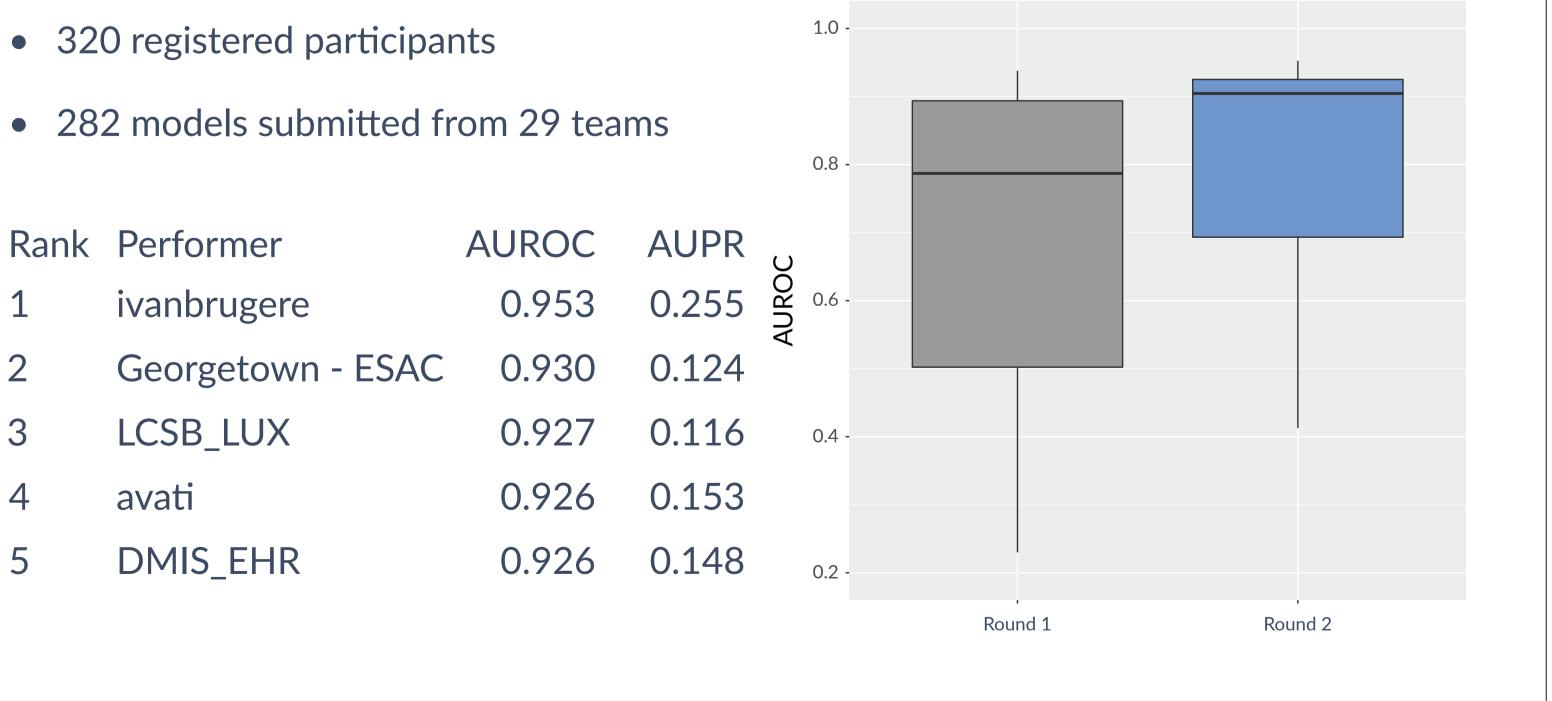




## **Participation & preliminary results**

- 282 models submitted from 29 teams

#### Evolution of the AUROC of submitted models





Clinical data

(longitudinal)

recent exam

• Participants do no have **direct access to patient data** at any point.

Digital mammograms

(longitudinal)







Dockerized model



<sup>\*</sup> Are contributing equally

- <sup>1</sup> Sage Bionetwork, Seattle, WA 98121, USA
- <sup>2</sup> Biomedical Informatics and Medical Education, University of Washington
- <sup>3</sup> Molecular Engineering & Sciences Institute, University of Washington
- <sup>4</sup> Institute for Translational Health Sciences, University of Washington

justin.guinney @sagebionetworks.org

**UW** Medicine

**BIOMEDICAL INFORMATICS** AND MEDICAL EDUCATION

## dreamchallenges.org

## synapse.org/ehr\_dream\_challenge\_mortality

Guinney, J., Saez-Rodriguez, J. Alternative models for sharing confidential biomedical data. Nat Biotechnol 36, 391–392 (2018), doi:10.1038/nbt.4128 Ellrott, K., Buchanan, A., Creason, A., Mason, M., Schaffter, T., Hoff, B., ... & Saez-Rodriguez, J. (2019). Reproducible biomedical benchmarking in the cloud: lessons from crowd-sourced data challenges. Genome biology, 20(1), 1-9. doi.org/10.1186/s13059-019-1794-0

