**Topic:**

* **Kidney (how many patienst do we have), I need the numbers, atleast one function, self reported rate (200)**
* **For stroke (check too)**

**What is the best strategy risk prediction model for diseases, should we use race indicator as a predictor in the model. What are the variable less likely to make the model less transferaböle. What population can I use the model for.**

**Whereather there are differences across differences,**

**Transportatpility, how does it differ across all the diseases.**

**Transfer learing, can the new advancements solve the new problem?**

**Including social determinnats of health or not? Include race or not? Treatment effect across groups**

**Knowledge graph**

**How does that work between different groups.**

**Causal Inference :**

**wHAT IS THE POPULATION DECOMPOSITION, WHAT ARE THE FACTORS THAT AFFECT THESE MOST?**

**Investigate whether certain treatments have different effect size across different sub groups.**

* Polygenic risk scores

Gap between method and reality: data limitations,

Different training strategies, pros and cons, how how we are still missing for a certain disease, or group of diseases,

Comnbiniation of age, social determinants of health

First Save genetic and social determinants of health, then.

Differences across different diseases: stroke, type 2 diabetes.

Goal for next weeks:

* Explore data,
* Prepare workspace
* Sample size
* Add background
* Models: Penalized regression (Lasso, ridge), XGBOST, rANDOM FOREST (DECISION TREE), BART, Data driven vs non risk factors

**Research Topic 1: Evaluating the Impact of Social Determinants of Health on Risk Prediction Models for Kidney Disease and Stroke**

* **Objective**: To develop and compare risk prediction models for kidney disease and stroke, incorporating traditional clinical risk factors and social determinants of health (SDH), such as socioeconomic status, education, and race/ethnicity.
* **Methods**: Use a large dataset incorporating both clinical and SDH factors. Employ advanced statistical models like penalized regression and machine learning algorithms (e.g., XGBoost and Random Forest) to assess the predictive accuracy and fairness of each model.
* **Impact**: This research could inform personalized treatment plans and public health strategies, highlighting the importance of considering SDH in clinical risk assessments.

**Research Topic 2: Cross-Disease Analysis of Treatment Effectiveness Using Causal Inference and Polygenic Risk Scores**

* **Objective**: Investigate the variability in treatment outcomes across different diseases (kidney disease, stroke, type 2 diabetes) by employing causal inference methods and integrating polygenic risk scores.
* **Methods**: Apply methods like BART and other causal inference frameworks to evaluate the effect size of treatments across subgroups defined by genetic and non-genetic risk factors. Use polygenic risk scores to stratify patients by genetic susceptibility.
* **Impact**: This study aims to refine personalized medicine approaches by identifying which subgroups benefit most from specific treatments, potentially leading to more tailored and effective therapeutic interventions.

**Research Topic 3: Assessing the Transferability and Transportability of Predictive Models Across Different Global Populations**

* **Objective**: To evaluate how well disease prediction models for kidney disease and stroke developed in one population perform when applied to different global populations.
* **Methods**: Develop risk prediction models using data from a primary population and assess their performance in secondary populations using statistical techniques for model transferability and transportability. Explore the role of race and other demographics in model performance to address potential bias.
* **Impact**: This research will address critical gaps in the generalizability of predictive models and could lead to more universally applicable public health strategies and clinical guidelines.

**Execution Plan**

* **Data Collection and Preparation**: Collect comprehensive datasets that include a wide range of demographics, clinical data, and social determinants of health. Prepare the data with rigorous preprocessing techniques.
* **Model Development and Validation**: Develop multiple models incorporating various predictors and compare their performance through cross-validation techniques and external validation on different datasets.
* **Statistical Analysis and Reporting**: Use advanced statistical analyses to interpret the data, and ensure the findings are reported transparently and robustly,