BIO SM60 Thesis Proposal

Student Name	Thesis Advisor Name
Ndey Isatou Jobe	Dr. Rui Duan
Proposed Research Project Title Evaluating Disparities in End-Stage Renal Disease Risk Prediction Using the All of Us Cohort	
Organization where the project will be completed (if not Harvard Chan School of Public Health)	
Proposed Start Date (mm/dd/yyy) May 15th 2024	Proposed End Date (mm/dd/yyy) December 15th 2024
Thesis Advisor's Approval	Date (mm/dd/yyy)
BIO SM Program Director's Approval	Date (mm/dd/yyy)

Expectations (to de decided with the thesis advisor). This should include frequency of meetings, attendance at group meetings, hours worked each week, etc.

Please describe your proposed research project, including the sections below. Your report should be 1-2 pages in length.

Introduction (Hypothesis Being Tested, Background, Significance)

Proposed Methods

Anticipated Results

Proposed Research Project

Title:

Evaluating Disparities in End-Stage Renal Disease Risk Prediction Using the All of Us Cohort

Introduction

Study Hypotheses and Objectives

Primary

This study seeks to develop and validate two- and five-year risk prediction models for the progression to End-Stage Renal Disease (ESRD), with a focus on the role of genetic ancestry as a significant determinant of ESRD risk.

Secondary

1. APOL1 Risk Variants: Apolipoprotein L1 (APOL1) risk variants, which are prevalent among individuals of African ancestry, are associated with an increased risk of ESRD [2][4][5]. This study will examine the impact of comorbid conditions and exposures on ESRD progression in patients with high-risk APOL1 alleles.
2. Proteinuria: Given the established link between proteinuria and chronic kidney disease (CKD) progression [8], this study aims to investigate its specific contribution to racial disparities in ESRD incidence.
3. Socioeconomic Factors: This study will assess the impact of socioeconomic status on racial disparities in ESRD, examining how differences in socioeconomic variables influence the risk and progression of ESRD.

Background

Significant racial and ethnic disparities in End-Stage Renal Disease (ESRD) incidence have been documented, with notable increases across various populations between 2000 and 2019 [6]. Despite these disparities, current ESRD risk calculators often omit race as a factor, and traditional risk factors alone do not fully account for the observed differences [2]. This study seeks to explore the role of genetic ancestry in ESRD progression by examining baseline estimated Glomerular Filtration Rate (eGFR) levels, with the goal of deepening our understanding of health inequities in kidney disease.
APOL1 risk variants, which are prevalent among individuals of African ancestry, are known to substantially increase the risk of ESRD. However, not all carriers of these variants develop kidney disease, indicating the presence of other contributing factors [2]. This study

aims to identify and characterize additional risk factors for ESRD, including potential modifiers of the APOL1 effect.
Proteinuria, a well-established risk factor for the progression of chronic kidney disease (CKD) to ESRD, may also play a significant role in the disparities observed in ESRD risk [2] [7]. This study will investigate the contribution of proteinuria to these disparities.
Furthermore, socioeconomic factors may substantially influence disparities in ESRD outcomes. This study will assess the impact of these variables on ESRD risk and progression, with the aim of elucidating their role in contributing to health inequities.

Significance

Recognizing and addressing racial disparities in the progression of end-stage renal disease (ESRD) is essential for crafting targeted interventions and shaping informed clinical and public health policies. Gaining these insights can foster more equitable healthcare practices, alleviate the burden of ESRD on vulnerable populations, and ultimately enhance patient outcomes.

Proposed Methods

This retrospective cohort study will utilize data from the "All of Us Controlled Tier Dataset v7." The study cohort will consist of 413,457 participants aged 18 years and older from the All of Us Research Program.

Statistical Analysis

Part 1: Prediction Models

Machine Learning Algorithms

This study will employ several approaches to develop and evaluate risk prediction models for end-stage renal disease (ESRD):
1. Evaluation of Existing Risk Prediction Models: Assessing the Performance of the Standard ESRD Prediction Model [1] [3] [9] [10] on the All of Us Data for Two- and Five-Year Risk.
2. Training New Risk Prediction Models with Race as a Predictor: Develop new models incorporating race as a variable to determine its impact on predictive accuracy.
3. Training New Risk Prediction Models without Race as a Predictor: Develop models that exclude race to evaluate predictive performance without this variable.
4. Advanced Fairness-Aware Machine Learning Methods (Optional): Implement methods that address fairness and bias in predictive modeling to ensure equitable outcomes across different demographic groups (Optional).

5. Advanced Multi-Task Learning Methods (Optional): Explore multi-task learning approaches that can simultaneously predict multiple related outcomes, potentially enhancing model performance and utility (Optional).

Part 2: Contributions of genetic and socioeconomic status (SDoH) in risk prediction

This study aims to comprehensively assess the extent to which genetic risk factors, socioeconomic status, and environmental factors contribute to disparities in End-Stage Renal Disease (ESRD) risk.
1. APOL1 Risk Variants: We will utilize Inverse Probability Weighting (IPW) and Cox proportional hazards regression analysis to evaluate the impact of APOL1 risk variants on the progression to ESRD.
2. Proteinuria: The association between proteinuria levels and ESRD risk will be examined using multivariable logistic regression.
3. Socioeconomic Factors: The impact of socioeconomic factors on ESRD risk will be assessed using multivariable linear and logistic regression models.
All statistical analyses will be conducted using R (version 4.0.3).

Anticipated Results

In this retrospective study examining the role of genetic ancestry in predicting the progression of end-stage renal disease (ESRD), we anticipate that existing risk prediction models will demonstrate moderate accuracy but also highlight significant racial disparities. Incorporating race into new models is expected to improve predictive accuracy; however, this approach may also raise fairness concerns. Machine learning techniques may enhance the overall predictive power, yet persistent racial disparities are likely to remain evident. Fairness-aware and multi-task learning methods are projected to provide more equitable and comprehensive predictions.
Furthermore, APOL1 variants, proteinuria, and socioeconomic factors are anticipated to significantly influence ESRD risk, with notable differences observed across racial groups. This study aims to underscore the multifactorial nature of ESRD risk and to enhance the understanding and equitable prediction of racial disparities. By addressing these complexities, the research seeks to inform targeted interventions and promote more equitable healthcare outcomes.

References

[1] Herrington WG Aguilar-Ramirez D. In ckd, the kidney failure risk equation predicted 2-y risk for eskd better than egfr alone. *Ann Intern Med*, 175(5):JC59, 2022. PubMed ID: 35500267.

[2] Fabian Bock, Thomas G. Stewart, Cassianne Robinson-Cohen, Jennifer Morse, Edmond K. Kabagambe, Kerri L. Cavanaugh, and Kelly A. Birdwell et al. Racial disparities in end-stage renal disease in a high-risk population: The southern community cohort study. *BMC Nephrology*, 20(1), 2019.

[3] Anderson AH et al. Bundy JD, Mills KT. Prediction of end-stage kidney disease using estimated glomerular filtration rate with and without race: A prospective cohort study. *Ann Intern Med*, 2022. PubMed ID: 35007146.

[4] Carlos Eduardo Duran, Mayra Estacio, Daniela Espinosa, Eliana Manzi, Juan G. Posada, Liliana Mesa, and Johanna Schweineberg. Apol1 gene variants and risk for cardiovascular disease. *Kidney and Blood Pressure Research*, 48(1):785–790, 2023.

[5] Ai Itoku, Jaya Isaac, Scott Wilson, Kimberly Reidy, and Frederick Kaskel. Apol1 nephropathy risk variants through the life course: A review. *American Journal of Kidney Diseases*, 84(1):102–110, 07 2024.

[6] Bridget M. Kuehn. End-stage kidney disease doubles. *JAMA*, 327(16):1540, 2022.

[7] Zhalaliddin Makhammajanov, Abduzhappar Gaipov, Askhat Myngbay, Rostislav Bukasov, Mohamad Aljofan, and Mehmet Kanbay. Tubular toxicity of proteinuria and the progression of chronic kidney disease. *Nephrology Dialysis Transplantation*, 39(4):589–599, 10 2023.

[8] David Pitcher, Fiona Braddon, Bruce Hendry, Alex Mercer, Kate Osmaston, Moin A. Saleem, Retha Steenkamp, Katie Wong, A. Neil Turner, Kaijun Wang, Daniel P. Gale, and Jonathan Barratt. Long-term outcomes in iga nephropathy. *Clinical Journal of the American Society of Nephrology*, 18(6):727–738, 04 2023.

[9] Griffith J et al. Tangri N, Stevens LA. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA*, 305(15):1553–9, 2011. PubMed ID: 21482743.

[10] Levey AS et al. Tangri N, Grams ME. Multinational assessment of accuracy of equations for predicting risk of kidney failure: A meta-analysis. *JAMA*, 315(2):164–74, 2016. PubMed ID: 26757465.