Development and Fairness Evaluation of CVD Risk Prediction Models for Individuals with Type 2 Diabetes

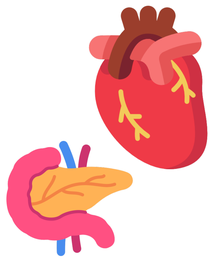
An AIM-AHEAD CDP Project

Yang Yang, BS[1], Tian Liu, MS[2], Che-Yi Liao, MS[1], Sun Ju Lee, MS[1], Esmaeil Keyvanshokooh, PhD[2], Hui Shao, PhD[3], Mary Beth Weber, PhD[3], Francisco Pasquel, MD, MPH[3,4], Gian-Gabriel P. Garcia, PhD[1]

[1]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Tech, Atlanta, GA; [2]Mays School of Business, Texas A&M University, College Station, TX; [3]Rollins School of Public Health, Emory University, Atlanta, GA; [4]Emory University School of Medicine, Atlanta, GA

BACKGROUND

Type 2 Diabetes (T2D) is one of the most common chronic conditions in the USA, **affecting over 38 million Americans**



A significant proportion of people with T2D are at **increased risk of cardiovascular disease** (CVD), i.e., heart failure, stroke, and myocardial infarction (heart attack)

To facilitate CVD risk management, several CVD risk equations such as the PCEs and PREVENT have been developed.

Yet, many such risk equations were:
- developed on data that **do not reflect the diversity of the US** population and
- have not been evaluated for **predictive fairness**

The **NIH All of Us** dataset, which over-represents minority populations and contains SDOH variables, **can facilitate the fair development and evaluation of risk equations**, but has yet to be applied to CVD risk estimation for people with T2D.

Research Objectives

By leveraging the NIH All of Us dataset, our objectives were to:
1. develop a machine learning (ML) model to estimate CVD risk among people with T2D
2. compare the ML model's accuracy and fairness to PREVENT across subgroups by sex and race/ethnicity

METHODOLOGY

All of Us RESEARCH PROGRAM → Data Pre-processing → Train 80% N=19k / Test 20% N=4.7k

We modeled the time to a CVD event using the **Weibull AFT** model:

$$\lambda(t|\theta) = \theta\lambda_0(\theta t); \quad \theta = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

Performance metrics:
- Variable importance
- Model calibration
- Predictive accuracy via **concordance** (C-index)
- Fairness (concordance imparity, concordance fraction)

Both the **Weibull AFT** and **PREVENT** were evaluated on the test set

RESULTS

Training Data N=19,036	Testing Data N=4,759
Sex:	Sex:
• Female: 56.51%	• Female: 55.83%
• Male: 41.16%	• Male: 41. 61%
• Other: 3.00%	• Other: 2.56%
Race/Ethnicity	Race/Ethnicity
• NH Black: 27.14%	• NH Black: 26.39%
• NH White: 47.70%	• NH White: 48.69%
• Hispanic: 18.64%	• Hispanic: 18.83%
• Other: 6.51%	• Other: 6.09%
BMI, kg/m^2: 33.52 (6.09)	BMI, kg/m^2: 33.53 (6.90)
SBP, mm Hg: 131.23 (14.95)	SBP, mm Hg: 131.09 (14.46)
DBP, mmHg: 76.10 (10.83)	DBP, mmHg: 76.11 (12.75)
HbA1c, %: 7.16 (1.48)	HbA1c, %: 7.16 (1.48)
CVD events: 28.19%	CVD events: 27.11%

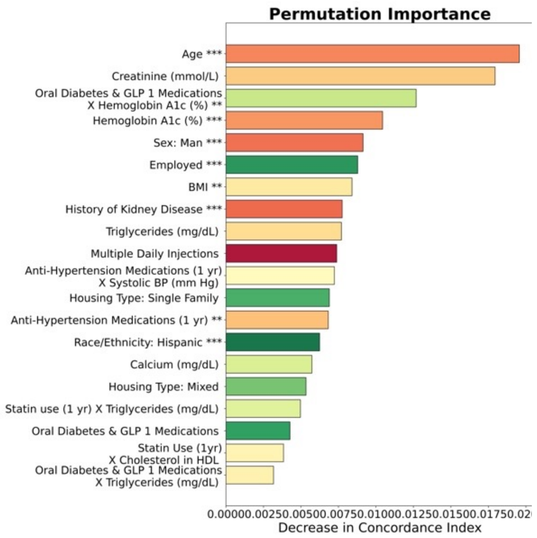


Figure 1. *Permutation variable importance. Beyond what is included in PREVENT, important variables in the Weibull AFT model include kidney function and blood glucose, medication history, demographics, and SES*

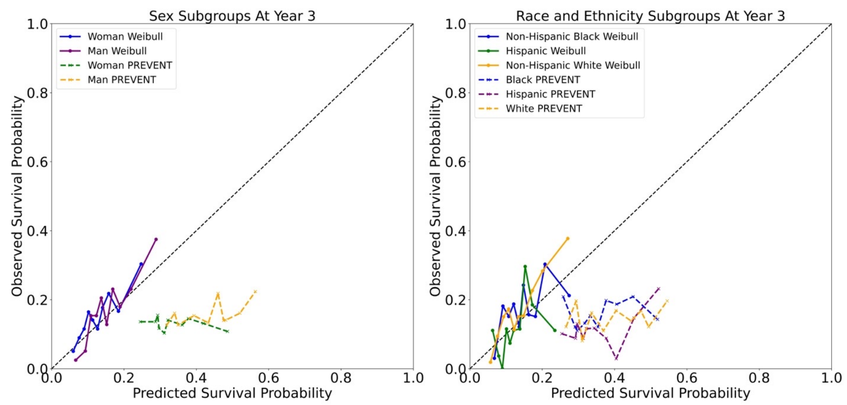
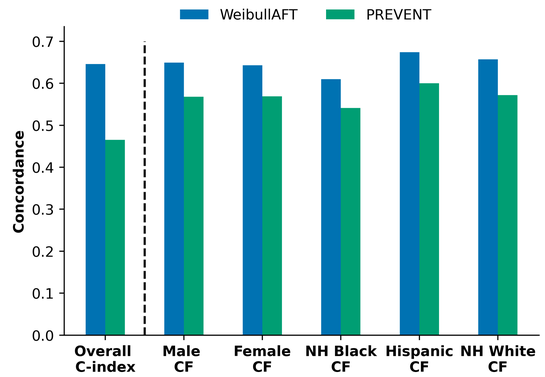


Figure 2. *Calibration plots. Weibull AFT is better calibrated than PREVENT, which often over-estimates 3-year CVD risk*

Figure 4. *Concordance. Weibull AFT achieves higher C-index overall and higher concordance fraction (CF) across all demographic groups*



IMPLICATIONS

On people with T2D, Weibull AFT is much more accurate (i.e., concordant) and better-calibrated than PREVENT

Overall performance for each sex and race/ethnicity subgroup is better with Weibull AFT, though PREVENT has lower Concordance Imparity

For patients with T2D, measures of kidney function and other T2D risk factors play an important role in predicting risk of CVD

Demographics and SES also play an important factor in predicting risk of CVD events

CONCLUSIONS

Patients with T2D require tailored algorithms for accurate risk estimation and treatment planning

Datasets such as NIH All of Us, which over-represent traditionally under-represented patient populations play an important role in fair ML

Holistic model development approaches are necessary for creating responsible ML and AI models

ACKNOWLEDGEMENTS & DISCLAIMER

This research is, in part, funded by the National Institutes of Health (NIH) AIM-AHEAD Program Agreement NO. 1OT2OD032581. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the NIH.