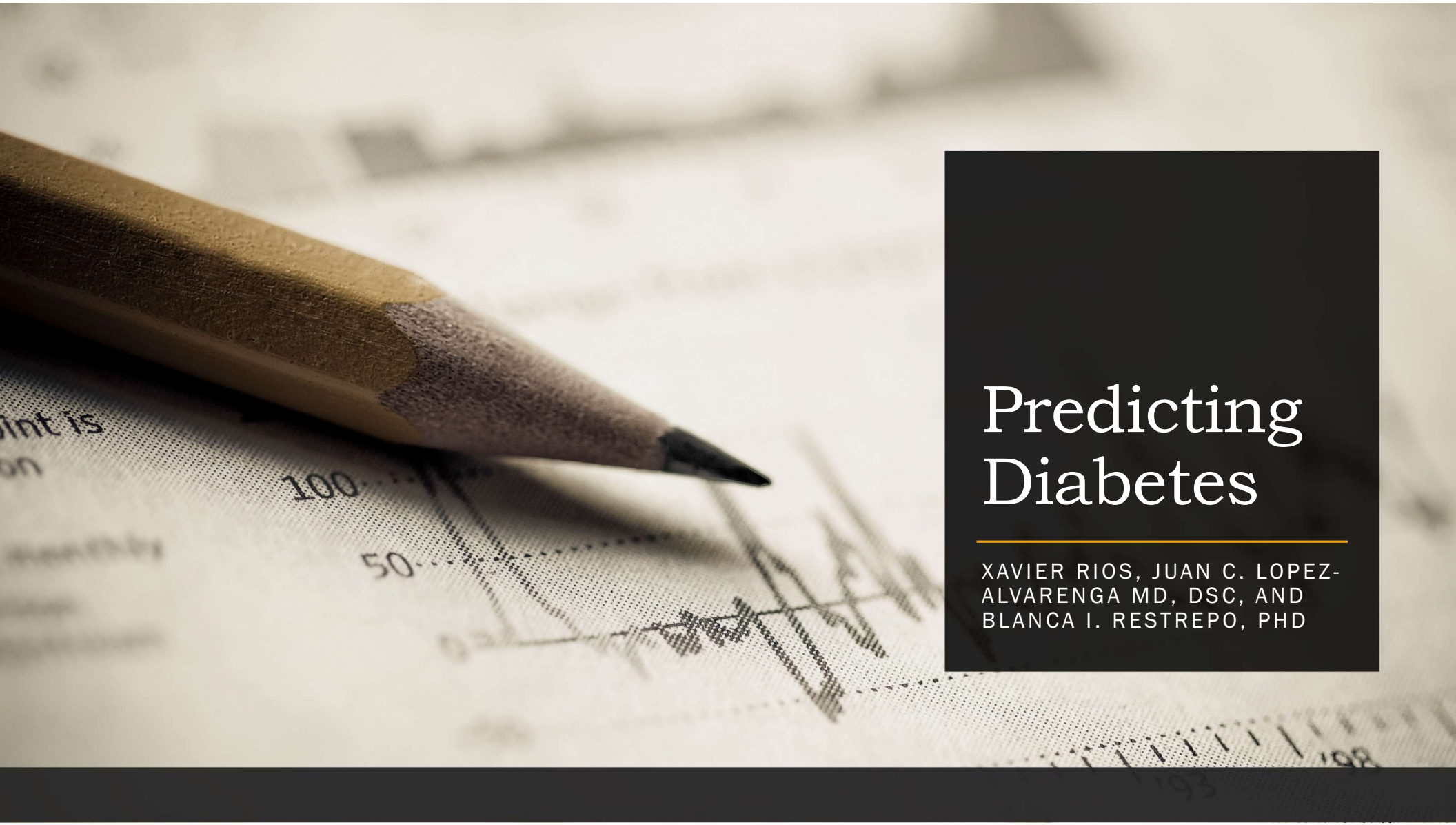


Predicting Diabetes

XAVIER RIOS, JUAN C. LOPEZ-ALVARENGA MD, DSC, AND BLANCA I. RESTREPO, PHD

Background

The active search for type 2 diabetes in clinics results in inefficient use of time and resources when using invasive laboratory exams (e.g. blood glucose, HbA1c).

Simple methods like anthropometry or clinical history could help prioritizing testing in individuals with higher risk of type 2 diabetes.

- That way, those who are tested have a higher probability of having diabetes.

Goal of our study:

Design models to predict type 2 diabetes using non-invasive clinical measurements.

Evaluate the predictive potential of these models.

Methods

Study Design: Cross-sectional nested design on NHANES 2017-2018 survey (9,254 participants)

1. Data Cleaning
   - Excluded participants:
     - Under 20 years and over 60 years old
     - With missing diagnostic or predicting variable data
   - Final result: 3,037 participants

2. Searched for variables to define our outcome of interest: Who has **Diabetes**?
   - Participants diagnosed following the American Diabetes Association (ADA) parameters
     - **Self-reported diabetes**, or
     - **Newly diagnosed diabetes** if:
       - HbA1c $\geq$ 6.5%, diabetes, or
       - Fasting blood glucose $\geq$ 126 mg/dL (fasting $\geq$ 6 hours since last meal), or
       - Non-fasting blood glucose $\geq$ 200 mg/dL (non fasting $<$ 6 hours since last meal)

3. Identified predicting variables (e.g. variables that we anticipate will influence a diabetes diagnoses)
   - Age
     - Included participants from 20 years to 60 years old
     - Grouped into 3 categories of age, each spanning 20 years

Methods Con't

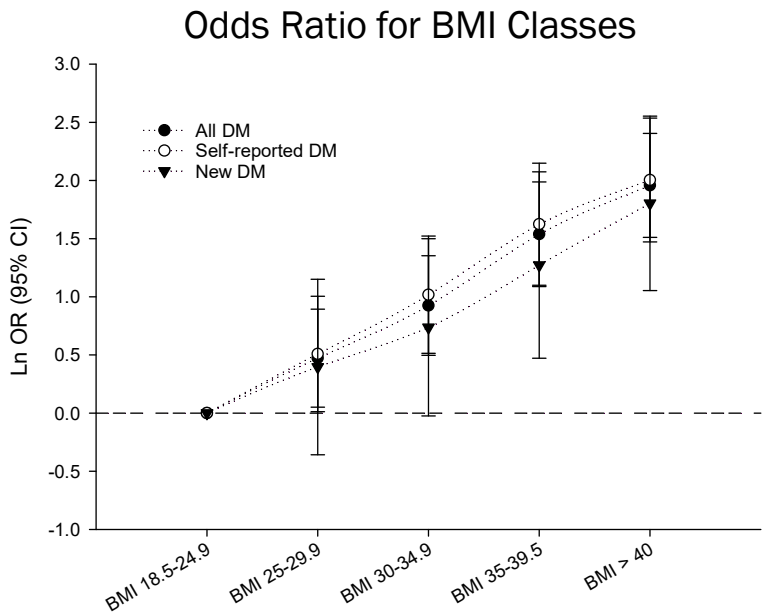
- Family History
  - Family history of diabetes was grouped into "No family history" and "Yes family history"
  - Participants who had self-reported diabetes were not asked this question; this most likely affected our odds ratio estimates as shown on a future slide
- BMI
  - Participants BMI were stratified into 5 classes
    - Normal Weight – 18.5-24.9
    - Overweight – 25-29.9
    - Obesity Class 1 – 30-34.9
    - Obesity Class 2 – 35-39.5
    - Obesity Class 3 - > 40

4. Variables for prediction analyzed, weighted, and combined for ideal model

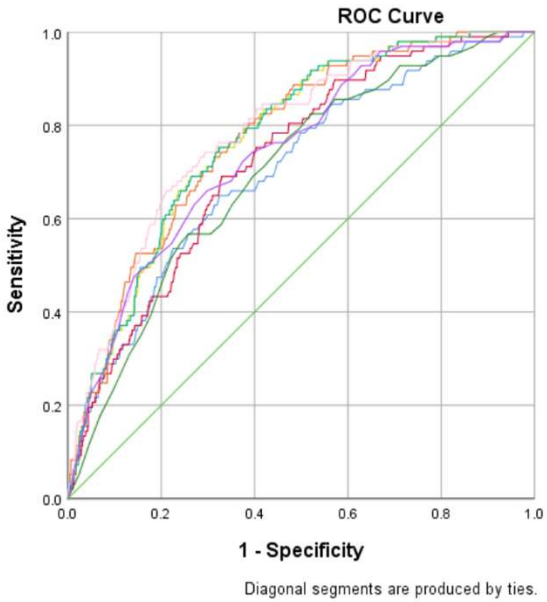
- ANOVA
- Odds ratios (95% CI)
- Chi-square
- Logistic regression
- ROC curves

Odds Ratio of DM Diagnosis

	Odds for DM Diagnosis		
BMI Range (18.5-24.9 as base)	**All DM**	**Self-Reported DM**	**New DM**
25-29.9	1.604	1.663	1.487
30-34.9	2.522	2.768	2.091
35-39.5	4.657	5.075	3.572
>40	7.086	7.422	6.075
Age (20-40 as base)	**All DM**	**Self-Reported DM**	**New DM**
>40	5.785	7.045	3.829
Family History (No FH as base)	**All DM**	**Self-Reported DM**	**New DM**
Yes	0.331	0	1.765



Comparison of Different Models



Combined ROC Curve of Models 1-8 with reference line

Source of the Curve

- Body Mass Index (kg/m**2)
- Waist Circumference (cm)
- Age in years at screening
- BMI_AGE
- BMI_AGE_WAIST
- BMI_AGE_WAIST_FAM
- BMI_AGE_FAM
- WAIST_AGE_FAM
- Reference Line

Area Under the Curve

	Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
					Lower Bound	Upper Bound
Model 1	Body Mass Index (kg/m**2)	.703	.027	.000	.651	.756
Model 2	Waist Circumference (cm)	.724	.024	.000	.678	.771
Model 3	Age in years at screening	.698	.025	.000	.649	.748
Model 4	BMI_AGE	.775	.021	.000	.733	.817
Model 5	BMI_AGE_WAIST	.775	.021	.000	.734	.816
Model 6	BMI_AGE_WAIST_FAM	.775	.021	.000	.734	.816
Model 7	BMI_AGE_FAM	.783	.022	.000	.740	.826
Model 8	WAIST_AGE_FAM	.742	.024	.000	.694	.790

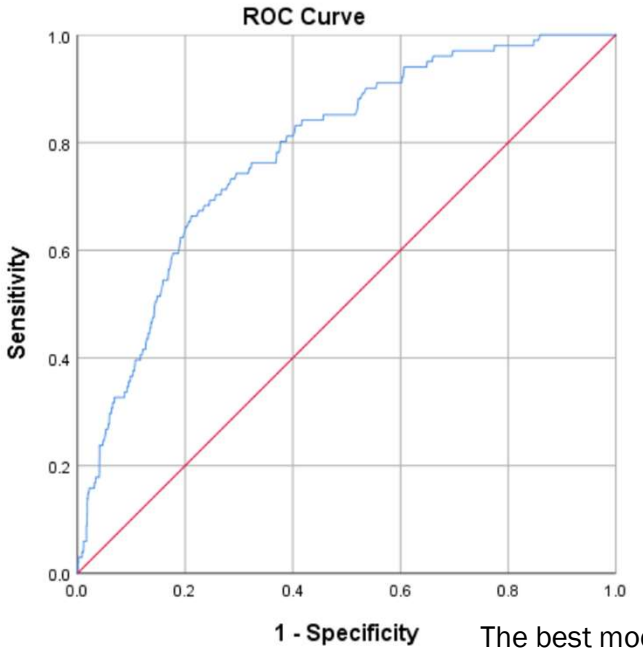
The test result variable(s): Body Mass Index (kg/m**2), Waist Circumference (cm), Age in years at screening, BMI_AGE, BMI_AGE_WAIST, BMI_AGE_WAIST_FAM, BMI_AGE_FAM, WAIST_AGE_FAM has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

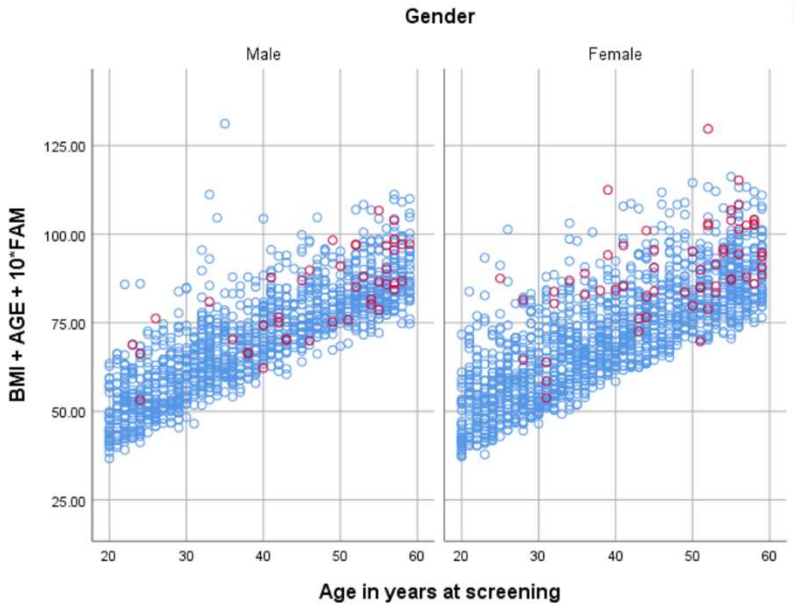
a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Notice the differing area under the curve (AUC). Model 7 proved to be the best predictor of type 2 diabetes.

Ideal Model for Prediction of New Type 2 Diabetes

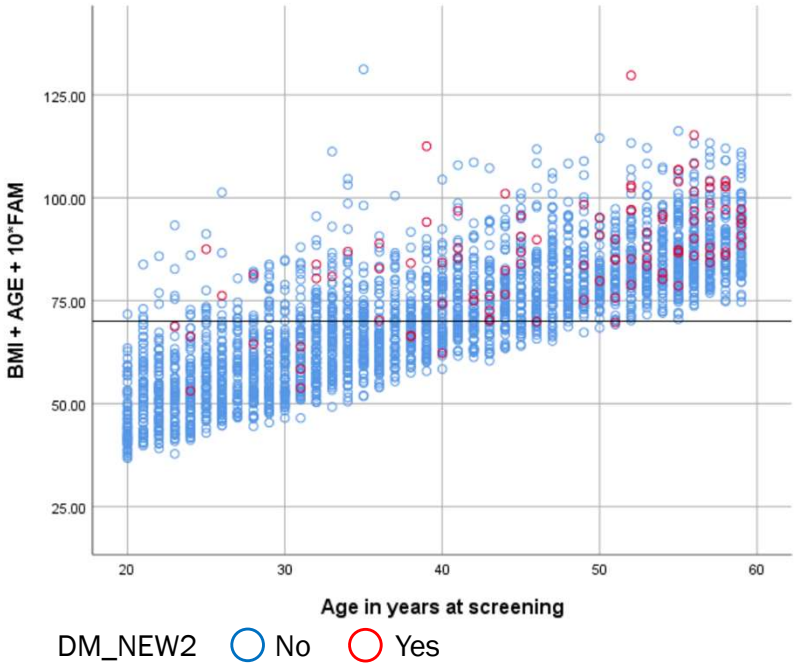




Red circles represent newly diagnosed participants with model score on y-axis

The best model (highest AUC = 0.783), included BMI, Age, and family history of DM

Sum of Age, BMI, and 10 for FHD scored at 70, Explanation



In our model, a score of 70 would be ideal for most clinics to go forward with DM testing:

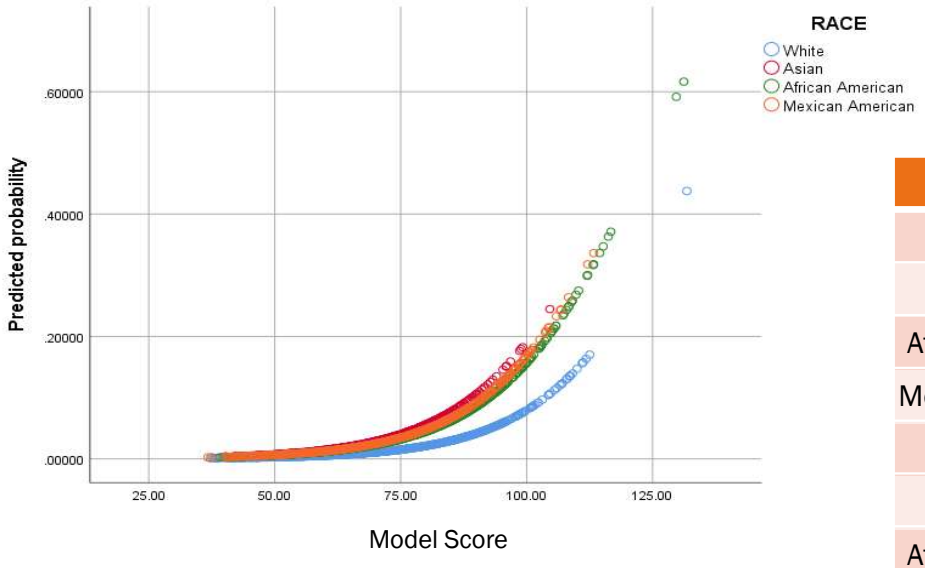
- ~90% sensitivity (~90% of new diabetics will be identified using the model)
- ~50% specificity (~50% of those tested will not have diabetes)

Clinics with more funding could test all patients, identifying everyone with DM, but using a lot of resources.

Clinics in under funded areas may have to test with a score above 70, identifying less DM patients, but most of those tested will have diabetes.

Using Race and Ethnicity to Determine Predictability

Scatter Plot of Model Score vs. Predicted Probabilities

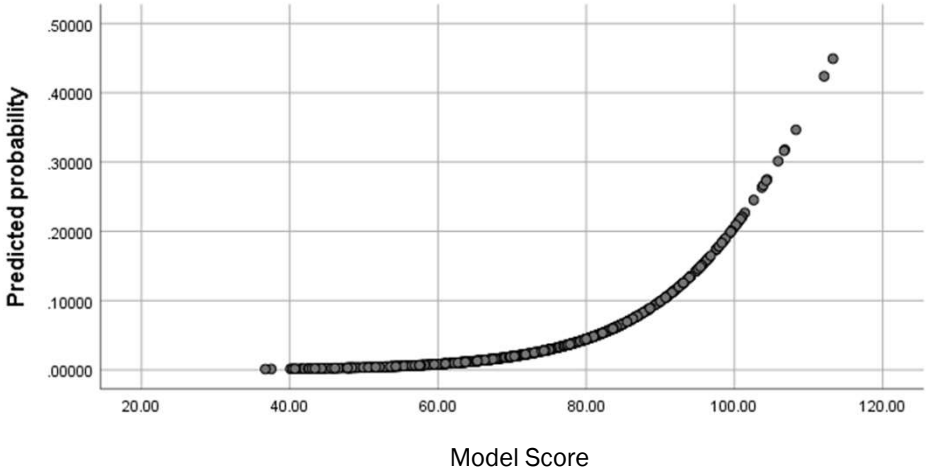


Found that minorities were about twice as likely to be detected with new diabetes using the model than White participants.

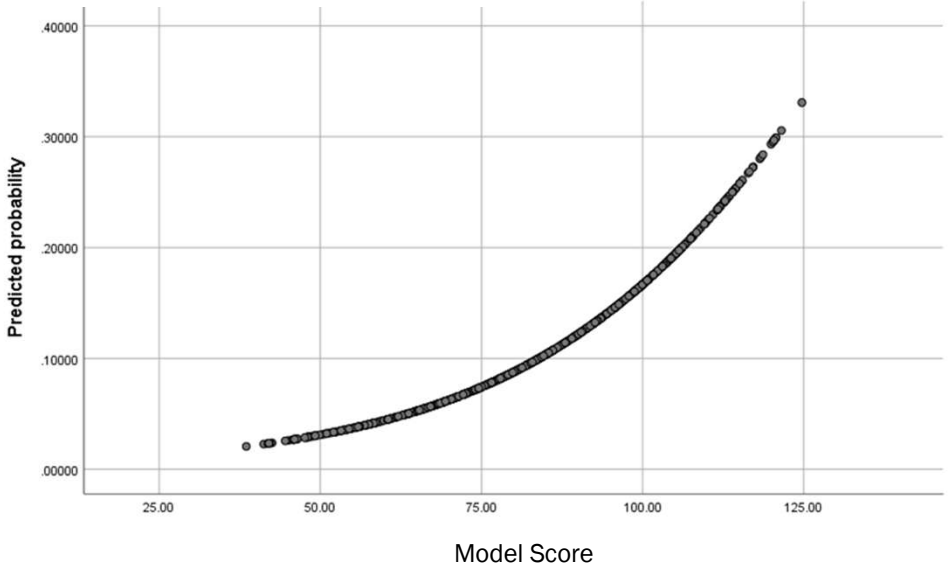
Race	Score	Probability	Probability Ratio (White as reference)
White	70	0.0107	1.0000
Asian	70	0.0287	2.6769
African American	70	0.0228	2.1249
Mexican American	70	0.0247	2.3020
White	120	0.2545	1.0000
Asian	120	0.4820	1.8943
African American	120	0.4234	1.6638
Mexican American	120	0.4435	1.7429

Comparing Predictability in NHANES vs. RGV Cohort

Scatter Plots of Model Score vs. Predicted Probabilities



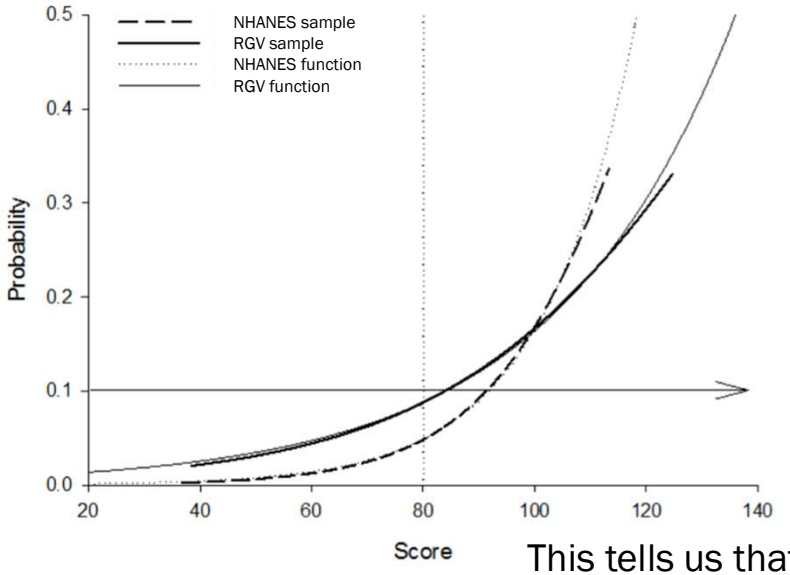
Filtered NHANES Cohort (Hispanic)



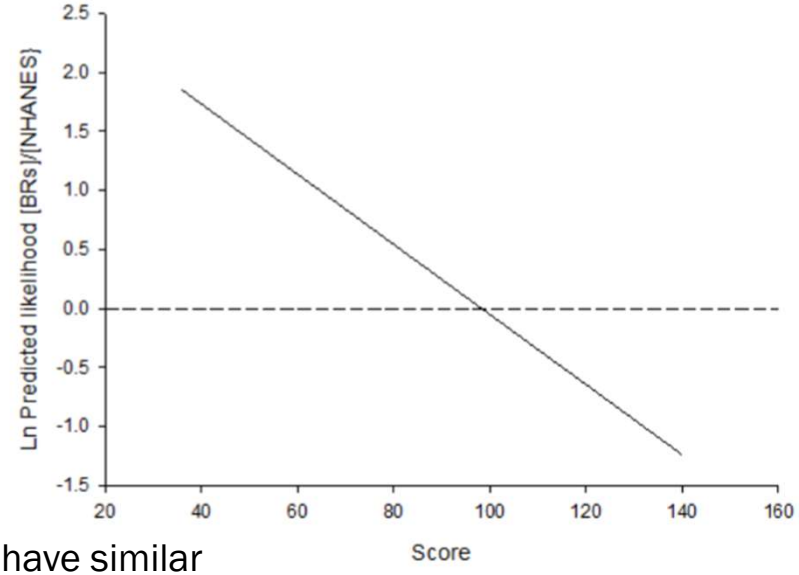
RGV Cohort (Mostly Hispanic)

Comparing Predictability in NHANES vs. RGV Cohort Con't

NHANES: $R^2 = 0.9972$, $f(x) = 0.0004 * e^{0.0607x}$
RGV cohort: $R^2 = 0.998$, $f(x) = 0.0074 * e^{0.0310x}$



The ratio RGV/NHANES shows important differences in the initial part of the function



This tells us that the models have similar predicting potential around a score of 100±20.

Accuracy and Applications of Model in RGV Cohort

Filtered composite model scores using 70 and 80 as testing boundaries:

- 70 as testing boundary
  - For those who were not previously diagnosed with diabetes, from our group of 586 participants
    - 409 would be tested
      - 56 would test positive for diabetes
      - 3 diabetic patients would not be tested
  - A clinic testing with a score of 70 would save 30% on costs versus sending everyone to testing.
- 80 as testing boundary
  - For those who were not previously diagnosed with diabetes, from our group of 586 participants
    - 290 would be tested
      - 46 would test positive for diabetes
      - 13 diabetic patients would not be tested
  - A clinic testing with a score of 80 would save 50% on costs versus sending everyone to testing.

Conclusions

We have identified a model, which had reasonable diagnostic accuracy for T2D using non-invasive measurements.

Based on the results from different model score boundaries, we have shown that a score of 70 would catch around 95% of diabetic patients in the RGV. Our model would require less resources for testing, making it more likely to be implemented in clinics.

In depth cost analysis is currently underway to determine the quality-adjusted life years gained with different composite model scores.