PREDICTIVE MODEL FOR CALCULATING INDIVIDUAL'S RISK OF TYPE 2 DIABETES

Mondal, Annika, Truong Chau, Ph.D.
UTHealth School of Public Health, Houston, TX, USA

BACKGROUND

Type 2 diabetes (T2D) is a metabolic disorder that occurs when the body's metabolism is disrupted or not functioning properly, often due to a chemical dysfunction. In T2D, the body continues to produce insulin but cannot metabolize it effectively. According to the CDC, type 2 diabetes (T2D) is prevalent among adults in the United States. However, more children and teenagers have developed this disorder recently. There has been an increase in the number of overweight children, which is closely associated with the development of T2D due to insulin resistance. T2D risk factors include but are not limited to a poor diet, a sedentary lifestyle, family members who have the disease, high cholesterol, and high blood pressure. Based on recent literature, these covariates have been used previously to predict the risk of undiagnosed diabetes (Collins et al., 2011).

SIGNIFICANCE

This research aims to analyze the health records within IBM Explorys EMR data to identify people diagnosed with T2D and their risk factors and then develop a predictive model incorporating the identified risk factors. The predictive model can quantify the likelihood of an individual with type 2 diabetes (T2D) based on their historical medical test results and current health status. Although genetic factors may increase a child's risk of having T2D, such as race and family medical history of having T2D or gestational diabetes, there are some ways for them to control or prevent T2D. Such methods would include exercising and having a balanced diet. Early detection of an individual's risk of having T2D would help the individual avoid severe complications associated with T2D (i.e., vision loss, kidney disease) (CDC, 2021a; CDC, 2021b). This project will also identify generalized steps to reduce the risk of T2D. The future application of this model will create awareness in the community and be an early detection tool for the study population.
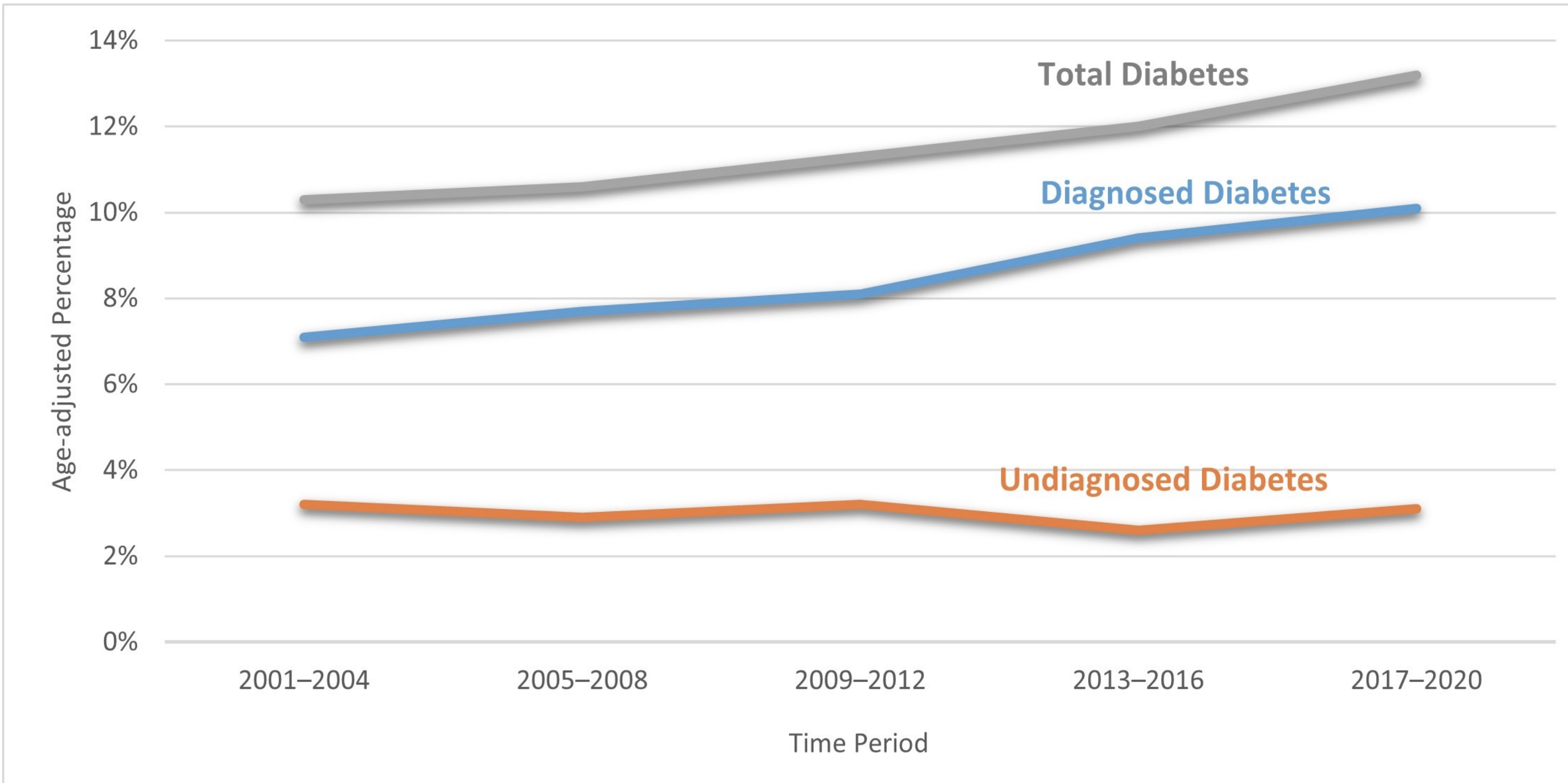

FIGURE 1

Figure adapted from CDC's National Diabetes Statistics Report. The total trend in the number of diabetes cases is rising for adults 18 years or older.

DATA USED

This study will use IBM Explorys EMR data to create a predictive model to calculate an individual's risk of type 2 diabetes. The Explorys EMR data is fully de-identified and comprises EMR data submitted by participating providers nationwide. The data set includes nearly 70 million unique individuals and inpatient, outpatient, and professional records linked by an encrypted, unique person ID. The study poses no risk to privacy as it is fully de-identified. Multivariable logistic regression was used to derive the predicted model. Upon literature review, we included variables clinically associated with Type 2 diabetes that could be extracted from Explorys. Age, gender, and race are demographic variables that can influence an individual's likelihood of having a Type 2 diabetes diagnosis. Age was transformed from a continuous variable into a categorical variable. Phypeglycemia, Hypertension, low-density lipoprotein, and hyperglycemia are medical conditions that have been shown to be strongly associated with Type 2 diabetes. These variables were utilized as binary variables in the model. If an individual had a diagnosis code for a condition, the variable had a flag of one, or else the variable had a flag of zero. These variables were then fitted into a predicted model to obtain regression coefficients which were then utilized to predict the probability of having Type 2 diabetes based on a patient's demographic and medical characteristics.

METHODS

Descriptive analysis will be employed to provide an overall look at T2D and covariates with the study population. Chi-square and t-test will be used to create p-values to determine the statistical significance between variables. This study will use a modified method seen in Ranapurwala et al.'s study (2019) to create a predictive model. Multivariable logistic regression will be employed to produce odds ratios at an individual level. After model fit has been determined using AIC, the estimated model coefficients from the test model will be used on the validation dataset. Finally, predictive probabilities will be calculated to determine the predicted probability of having T2D.

RESULTS

Covariates	Estimate	P-value
Intercept	-5.296	<.0001
Age Group		
0-17	-1.740	<.0001
35-44	0.786	<.0001
45-54	1.475	<.0001
55-64	2.028	<.0001
65-84	2.456	<.0001
Gender		
Male	Reference	
Female	-0.215	<.0001
Race		
White	Reference	
Black	0.457	<.0001
Asian	0.655	<.0001
Hispanic	0.669	<.0001
Unknown	0.369	<.0001
BMI		
Underweight (<18.5)	0.496	<.0001
Optimal (18.5-24.9)	Reference	
Overweight (25-29.9)	0.281	<.0001
Class I Obesity (30-34.9)	0.678	<.0001
Class II Obesity (35-39.9)	1.012	<.0001
Class III Obesity (>40)	1.224	<.0001
Other Health Conditions		
Hypertension	1.657	<.0001
Metabolic Syndrome	0.590	<.0001
Hyperglycemia	0.612	<.0001
Low HDL	0.158	<.0001
Phypeglycemia	0.726	<.0001

REFERENCES

1. Alustiza, E., Perales, A., Mateo-Abad, M., Ozcoidi, I., Aizpuru, G., Albaina, O., & Vergara, I. (2021). Tackling risk factors for type 2 diabetes in adolescents: PRE-STARt study in Euskadi. Anales de Pediatría (English Edition), 95(3), 186-196.
2. Centers for Disease Control and Prevention [CDC]. (2021, August 10). Prevent type 2 diabetes in Kids. Retrieved July 19, 2022, from https://www.cdc.gov/diabetes/prevent-type-2/type-2-kids.html
3. Centers for Disease Control and Prevention. (2021, December 16). Type 2 diabetes. Retrieved July 19, 2022, from https://www.cdc.gov/diabetes/basics/type2.html
4. Centers for Disease Control and Prevention. (2022, February 18). A SAS Program for the 2000 CDC Growth Charts (ages 0 to <20 years). Retrieved July 20, 2022, from https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm
5. Collins, G. S., Mallett, S., Omar, O., & Yu, L. M. (2011). Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC medicine, 9(1), 1-14.