2018

507 Final Project

TYPE 2 DIABETES

YAN, YUCHEN

Abstract

Type 2 diabetes is a long-term metabolic disorder that is characterized by high blood sugar, insulin resistance, and relative lack of insulin. Common symptoms include increased thirst, frequent urination, and unexplained weight loss. Type 2 diabetes is associated with a ten-year-shorter life expectancy. Therefore, with collected data of adult male of Type 2 diabetes patients, we analyzed the blood glucose level (mg/dL) association with race (white, black, Hispanic, or Asian), weight (kg), and hemoglobin A1C (%).

After analyzing the data collected by investigators, we conclude that the mean hemoglobin A1c level (%) differed across the four race groups ($p <.0001$), and that the pairwise of white and Black, white and Hispanic, Asian and Black, Asian and Hispanic are statistically significantly different ($p<.05$), and that the average of the group means for white and Hispanic patients is not significantly differed from the average of the group means for blacks and Asians ($p =.9775$). After using backward selection in regression model considering blood glucose levels (mg/dL) to be dependent variable, we include two independent variables: square of weight variable and hemoglobin A1C (%). This regression relationship differed across the four race groups with different intercept and different parameters respectively, but there is no modification evidence for race (all $p> .05$).

Introduction

Type 2 diabetes is a condition that impacts the way that the human body metabolizes sugar (glucose). It can become a chronic condition that affects the health of blood vessels and nerves, potentially leading to severe morbidity if not properly controlled. Investigators identified 491 adult male diabetics at the time of their first visit to the diabetes clinic at a large urban teaching hospital.

We sought to answer the following specific questions: first off, whether the mean hemoglobin A1C (%) differed across the four race groups. If so, we desire to know all possible pairwise race/ethnicity group comparisons of hemoglobin A1C (%). After examining the data for initial trends, we would test whether the average of the group means for white and Hispanic patients differed significantly from the average of the group means for blacks and Asians. Second, we were interested in the regression of blood glucose levels (in mg/dL) versus weight (kg) and hemoglobin A1C (%) (%). In particular, we would characterize whether and how this regression relationship differed across the four race/ethnicity groups.

For study purposes, we use collected data on a few key variables known to be potentially important in the etiology of diabetes. Data collectors identified 491 adult male diabetics at the time of their first visit to the diabetes clinic at a large urban teaching hospital. For study purposes, they collected data on a few key variables known to be potentially important in the etiology of diabetes. Specifically, the race/ethnicity (white, black, Hispanic, or Asian) of each patient was recorded at the initial visit along with his weight (kg), a random blood glucose measurement (mg/dL), and a measurement of hemoglobin A1c (%).

Method

With respect to whether the mean Hemoglobin A1C (%) across the four race/ethnicity groups, I use one-way ANOVA method to show difference between mean Hemoglobin A1C (%) across these four groups, and with check homogeneity of variances among four groups by adding “hovtest” in SAS. In this case, we should investigate all possible pairwise race group comparisons of hemoglobin A1C (%) and Scheffé's method tends to give narrower confidence limits and is therefore the preferred method. Then, because we are interested in testing whether the average of the group means for white and Hispanic patients differed significantly from the average of the group means for blacks and Asians, we add “estimate” line in SAS to request the estimate and SE for the contrast that we are interested in.

In selecting proper model for dependent variable blood glucose levels (mg/dL), I use backward selection among possible independent variables: hemoglobin A1C (%), weight (kg), hemoglobin A1C multiply weight, squared weight, squared hemoglobin A1C. After selecting proper model, we make overall regression model and regression model by four race, and then we use proc glm to estimate interaction by race with two independent variables. All tests performed were 2-tailed in our study. Our chosen significance level is 0.05 throughout the whole study and a p-value less than 0.05 indicates statistical significance in our study.

Results

From table 1, with total number of the patient number of white, black, Hispanic, and Asian are 123,135,109,124, and their mean of hemoglobin A1C (%) are 7.60, 7.89, 7.78, and 7.25 with standard deviation of 0.58, 0.68, 0.69, and 0.66, respectively. Also, we can see that the p-value of testing homogeneity of hemoglobin A1C (%) variance is 0.32>0.05, we accept null hypothesis and conclude that the hemoglobin A1C (%) have same variances among four race group. Therefore we can apply ANOVA method. From table 1, we can conclude that the mean hemoglobin A1C (%) differed across the four race groups (F value = 1.18, p < .0001).

Also, from table 1, the pairwise of white and Black, white and Hispanic, Asian and Black, Asian and Hispanic are significantly different (p<.05) and on the contrary, the pairwise of white and Asian, Black and Hispanic are not significantly different (p >.05). The average of the group means for white and Hispanic patients is not differed significantly from the average of the group means for blacks and Asians (p = .9775).

To look for proper regression model for blood glucose levels (in mg/dL) versus weight (kg) and Hemoglobin A1C (%) (%), from table 2, the summary of backward selection procedure in SAS, we can see that step 1 removed the variable Weight (kg)*hemoglobin A1C (%) with p value =0.68(F value =0.17); step 2 removed the variable Weight (kg)*hemoglobin A1C (%) with p value =0.62 (F value =0.24); step 3 removed the variable Square of hemoglobin A1C (%) with p value =0.32 (F value =0.98); finally, the variable of hemoglobin A1C (%) and the variable of Square of weight (kg) remained.

Therefore, the overall regression model can be expressed as:

$$GLUC = -36.69 + 0.001 * weight^2 + 22.72\, HBA1C$$

The white group regression model can be expressed as:

$$GLUC = -41.36 + 0.001 * weight^2 + 23.81\, HBA1C$$

The black group regression model can be expressed as:

$$GLUC = -71.51 + 0.002 * weight^2 + 26.45\, HBA1C$$

The Hispanic group regression model can be expressed as:

$$GLUC = -40.00 + 0.002 * weight^2 + 22.70\, HBA1C$$

The Asian group regression model can be expressed as:

$$GLUC = -41.98 + 0.002 * weight^2 + 23.742\, HBA1C$$

Where GLUC represents blood glucose levels (in mg/dL), the weight^2means variable of squared weight (kg) and the hemoglobin A1C (%) means variable hemoglobin A1C (%). All parameters in these models are valid ($p <.05$). From Table 4, the p value of interaction on hemoglobin A1C (%) by race are 0.95, 0.32, 0.81 for white, black and Hispanic groups respectively; and p value of interaction on weight^2 by race are 0.43, 0.81, 0.93, for white, black and Hispanic groups respectively.

Discussion:

From the results, with the same variances, the conclusion is that mean of hemoglobin A1C (%) differed across the four race groups, and that the pairwise of white and Asian, Black and Hispanic are same. Because these two pairwise exist, so that it is easy to understand the conclusion that the average of the group means for white and

Hispanic patients is same as the average of the group means for blacks and Asians (p = .9775).

From the regression model in results, we could conclude that all the predictors in the model are significant, so it is a valid model. Except for the intercept predictor, there are two predictors in the model. In this case, our five model's estimate parameter of squared weight predictor are all positive, meaning blood glucose levels have a positive association with squared weight (kg), which is corresponded to our common sense that heavy people may have a higher chance of high blood glucose levels. The hemoglobin A1C (%) predictor has significantly positive association with blood glucose levels (in mg/dL) in all five models, and this is also corresponded to the fact that A1C test can be used to diagnose type 2 diabetes and prediabetes. The higher the glucose level in your bloodstream, the more glucose will attach to the hemoglobin. Since this is the one with the best results in backward selection, the overall regression model is confirmed as the final model of the linear regression analysis and therefore, we get regression model for four groups respectively. However, we can see that the black group's parameters (intercept and slope) are slight different with other race groups from Table 3, so there may be modification by race. After the analysis, we confirmed there is no interaction exist by race in the overall model.

We should also be aware that this is a simplified model from a relative small sample and we only considered the adult males' information here. A more reliable analysis should not just consider male patients and should include external and environmental factors such as geographic area, air quality, etc.

References:

[1]. Kleinbaum, D.G., Kupper, L.L., Muller, K.E., and Nizam, A. (1998). *Applied Regression Analysis and Multivariable Methods*, 3rd, 4th or 5th edition, Duxbury Press.

[2]. Kutner, M.H., Nachtsheim, C.J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models, 5th edition*. WCB McGraw-Hill/Irwin, Boston.

Appendix

Table 1. The relationship among race groups with their Mean of hemoglobin A1C (%).

Characteristics	Number	Mean of hemoglobin A1C (%) with 95% CI	F Value (or t value)	p-value
Variance Homogeneity	491		1.18	0.32
White	123	7.36(6.79, 7.93)		
Black	135	7.89(7.21, 8.57)	0.014	< 0.0001
Hispanic	109	7.79(7.11, 8.48)		
Asian	124	7.25(6.59, 7.91)		
White - Black		-0.53(-0.76, -0.31)		< 0.05
White - Hispanic		-0.43(-0.67, -0.19)		< 0.05
White - Asian		0.11(-0.12, 0.34)		>= 0.05
Asian - Black		-0.64(-0.88, -0.41)		< 0.05
Asian - Hispanic		-0.54(-0.78, -0.30)		< 0.05
Black - Hispanic		0.10(-0.13, 0.34)		>= 0.05
White & Hispanic – Asian & Black		0.0033(-0.1145, 0.1211)	0.03 (t value)	0.98

Table 2. Backward selection summary.

Step	Removed Variable	F value	P value
1	Weight (kg)*hemoglobin A1C (%)	0.17	0.68
2	weight (kg)	0.24	0.62
3	Square of hemoglobin A1C (%)	0.98	0.32

* Finally, the variable hemoglobin A1C (%) and Square of weight (kg) remained.

Table 3. Overall regression model and regression model for four race group

Race group	Variable	Parameter Estimate	Standard Error	t value (or F value)	P value
White	Intercept	-41.36	17.58	-2.35	0.0203
	weight^2	0.001	0.0005	2.73	0.0072
	hemoglobin A1C (%)	23.81	2.44	9.75	<.0001
Black	Intercept	-71.51	14.84	-4.82	<.0001
	weight^2	0.002	0.0004	4.09	<.0001
	hemoglobin A1C (%)	26.45	1.90	13.91	<.0001
Hispanic	Intercept	-40.00	16.86	-2.37	0.0195
	weight^2	0.002	0.001	3.46	0.0008
	hemoglobin A1C (%)	22.70	2.20	10.33	<.0001
Asian	Intercept	-41.98	15.28	-2.75	0.0069
	weight^2	0.002	0.001	3.18	0.0018
	hemoglobin A1C (%)	23.42	2.14	10.96	<.0001
Overall	Intercept	-36.69	7.47	24.11(F value)	<.0001
	weight^2	0.001	0.0002	36.25(F value)	<.0001
	hemoglobin A1C (%)	22.72	1.02	497.17(F value)	<.0001

* The weight^2 means variable of squared weight (kg)

Table 4. Interaction analysis of squared weight (kg) and hemoglobin A1C (%) by race.

Parameter	Estimate	Standard Error	t Value	P value
hemoglobin A1C (%)				
hemoglobin A1C (%)*white	-0.19	3.15	-0.06	0.95
hemoglobin A1C (%)*black	2.82	2.82	1	0.32
hemoglobin A1C (%)*Hispanic	-0.73	2.96	-0.25	0.81
hemoglobin A1C (%)*Asian	0	.	.	.
weight^2				
weight^2*white	-0.0006	0.0008	-0.79	0.43
weight^2*black	-0.0002	0.0007	-0.24	0.81
weight^2*Hispanic	-0.0001	0.0008	-0.09	0.93
weight^2* Asian	0	.	.	.

* The weight^2means variable of squared weight (kg)
* All p value >0.05, there is no interaction by race for the two predictors