Original Article

The application of multivariate adaptive regression splines in exploring the influencing factors and predicting the prevalence of HbA1c improvement

Rui Lu[#], Tongqing Duan[#], Mengyang Wang[#], Hongwei Liu, Siyuan Feng, Xiaowen Gong, Hui Wang, Jiao Wang, Zhuang Cui, Yuanyuan Liu, Changping Li, Jun Ma

Department of Health Statistics, College of Public Health, Tianjin Medical University, Tianjin, China

Contributions: (I) Conception and design: C Li, J Ma, R Lu; (II) Administrative support: C Li, J Ma, Z Cui, Y Liu; (III) Provision of study materials or patients: C Li, J Ma, Z Cui, Y Liu; (IV) Collection and assembly of data: R Lu, T Duan, M Wang, H Liu, S Feng, X Gong, H Wang, J Wang; (V) Data analysis and interpretation: R Lu, T Duan, M Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Changping Li. No. 22 Qixiangtai Road, Heping District, Tianjin, China. Email: lichangping@tmu.edu.cn; Jun Ma. No. 22 Qixiangtai Road, Heping District, Tianjin, China. Email: junma@tmu.edu.cn.

Background: Glycosylated hemoglobin (HbA1c) is directly proportional to the level of glucose in the blood, and it has been the gold standard to evaluate the status of long-term blood glucose levels. Exploring the factors that lead to HbA1c improvement is beneficial for effectively controlling of HbA1c levels.

Methods: Data collected from 52 hospitals in five cities in northern China were divided into training and test sets at a ratio of 7:3. The training set was used to build models, and the test set was used to evaluate the generalizability of the models. The performance of multivariate adaptive regression splines (MARS) models and logistic regression was evaluated, namely, the accuracy, Youden's index, recall rate, G-mean and area under the ROC curve (AUC) with 95% confidence intervals (CIs).

Results: The prevalence of improvements in HbA1c levels was 38.35%. Doses of insulin less than 13 U, more than 3 kinds of oral medicine, exercise frequency greater than once per week and 2 h postprandial blood glucose (2hPBG) less than 10.56 mmol/L were found to improve HbA1c. The following interactions were negatively associated with improvement in HbA1c levels: patients with relative complications and 2hPBG less than 10.56 mmol/L, type 2 diabetes mellitus (T2DM) duration more than 7 years and insulin dose less than 13 U. Compared to logistic regression, the MARS model performed better in the above aspects, except for accuracy.

Conclusions: Given the interaction between factors affecting HbA1c improvement, medical staff should conduct comprehensive interventions to further reduce HbA1c levels in patients. In this study, the MARS model was superior to the traditional logistic regression in improving HbA1c levels. MARS had greater generalizability because it not only considered nonlinear relations in the process of model fitting but also adopted cross-validation. Nevertheless, more studies are needed to provide evidence for this result.

Keywords: Glycosylated hemoglobin (HbA1c) improvement; multivariate adaptive regression splines (MARS); logistic regression; model performance

Introduction

Type 2 diabetes mellitus (T2DM) is a common type of endocrine-metabolic disease. Clinically, it mainly manifests as hyperglycemia, and its serious complications lead to a decline in patients' quality of life, disability, and even death. Controlling blood glucose within a normal range is important to delay the development of these complications. Glycosylated hemoglobin (HbA1c) is directly proportional to the level of glucose in the blood, which can indirectly reflect the metabolic mechanism of glucose and compensate for the deficiency of traditional blood glucose monitoring. HbA1c has been considered the gold standard for evaluating the status of long-term blood glucose levels (1). Therefore, it is of great significance to explore the factors influencing HbA1c levels and predict the HbA1c improvement through these factors.

The classification problem is one of the important research topics in the field of data mining. Many existing classification methods are quite mature (2,3), and they can generally obtain good generalization performance. In previous studies, the most frequently used model for predicting the status of HbA1c was logistic regression (4-6). Logistic regression, one of the generalized linear model approaches, is a common classification method in data mining, especially for dichotomized classification problems. However, it rarely considers interactions between higher-order polynomials or risk factors. Therefore, logistic regression cannot analyze the nonlinear relationship between variables. In recent years, researchers have paid increasing attention to unconventional nonlinear modeling techniques. For example, multivariate adaptive regression splines (MARS) is an algorithm model that takes into account not only a linear relationship but also a nonlinear relationship, as well as the interaction between predictive variables. Therefore, it is of interest to explore the real relationship between the HbA1c improvement and related factors using the MARS model and to evaluate the prediction performance of the MARS model.

Methods

Subjects

This observational, cross-sectional study was conducted at 52 hospitals in five cities in northern China: Tianjin, Cangzhou, Tangshan, Qinhuangdao and Datong. Outpatients with T2DM who had begun basal insulin therapy for at least 3 months were consecutively recruited at each hospital from January 2015 to October 2017. We excluded (I) patients diagnosed with T1DM or gestational diabetes, (II) patients younger than 18 years old, (III) patients who were allergic to drugs, and (IV) patients with mental illness or poor compliance.

The information collected mainly included demographic characteristics and clinical information. Demographic information on age, gender, family history of T2DM, and physical activity was collected by trained interviewers through the completion of a questionnaire. Clinical information regarding height, weight, T2DM duration, insulin dose, oral medicine, related complications, HbA1c, 2 h postprandial blood glucose (2hPBG) and hypertension was obtained from field measurements and laboratory examinations. The height and weight of the enrolled participants were measured using standardized techniques, and their body mass index (BMI) was calculated. The HbA1c of all patients was measured within two weeks.

For most non-gestational patients with T2DM, an improvement in glycosylated hemoglobin (HbA1c) was defined as an HbA1c level at follow-up <7% (1,7). Therefore, patients in this study were divided into two groups at the cut-off point of 7%: an improvement group (HbA1c <7%) and a non-improvement group (HbA1c ≥7%).

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and approved by Medical Ethics Committee of Tianjin Medical University. All the subjects had signed the informed consent forms.

Theory

Multivariate adaptive regression splines
Multivariate adaptive regression splines (MARS) is a nonparametric method, introduced by Friedman (8), well suited to problems of high dimensional data. It divides the data space into several, possibly overlapping regions and fits truncated splines function in each region. Truncated splines functions consist of two linear splines functions, i.e., left sided Eq. [1] and right sided Eq. [2], separated from each other by a so-call knot (9).

$$b_q^-(x-t) = \left[-(x-t)\right]_+^q = \begin{cases} (t-x)^q \text{..............} if \ x<t \\ 0 \text{.........} otherwise \end{cases} \quad [1]$$

$$b_q^+(x-t) = \left[+(x-t)\right]_+^q = \begin{cases} (x-t)^q \text{.............} if \ x>t \\ 0 \text{.........} otherwise \end{cases} \quad [2]$$

where $b_q^-(x-t)$ and $b_q^+(x-t)$ are the splines functions describing the regions right and left of the knot location t, respectively, and q the power to which the splines is raised. The subscript "+" refers to the positive part. A splines function can also be called a basis function. For each of the explanatory variables MARS selects the pair of splines and the knot, that best describe the response variable. In the following step, the different basis functions are combined in one multidimensional model, which describes the response as a function of the explanatory variables. The result is a complex nonlinear model can be represented as:

$$\hat{y} = a_0 + \sum_{m=1}^{M} a_m B_m(x) \tag{3}$$

here $\hat{y}$ is the predicted value for the binary outcome; a_0, the coefficient of the constant basis function; M, the number of basis functions and B_m and a_m the *mth* basis function and its coefficient.

In general, a MARS analysis is carried out in three steps. In a first step independent variables explaining the response variable well are selected stepwise. Usually, the building global MARS model is very complex and shows overfitting. In the next step the model is pruned using iteratively a 10-fold cross validation and a general cross validation (GCV) procedure, resulting in an optimal MARS model. This is done using different sequences of Monte Carlo cross validation (10). Eventually, the model with the best accuracy is selected as the final output model in the third step, which is based on an evaluation of the predictive performance of the different models. For more information, see the refs (9,11).

Statistical analysis

Continuous variables, being normally distributed, are presented as the mean ± standard deviation, while categorical variables are expressed as counts and percentages. Comparisons between groups were performed using the independent Student's *t*-test for continuous data and Pearson's chi-square test for categorical data.

In this study, HbA1c, transformed into a binary outcome, was modeled as a response variable, and the demographic characteristics and clinical information mentioned above were used as explanatory variables. To compare the prediction performance of MARS and logistic regression, two methods were used to construct the models. All data were randomly divided into a training set and test set according to the ratio of 7:3. The training set was used to build models, and the test set was used to evaluate the

generalizability of the models. The performance of the two models was evaluated on the basis of accuracy, Youden's index, recall rate, G-mean and area under the ROC curve (AUC) with 95% confidence intervals (CIs).

All statistical analyses were carried out using R software version 3.5.1. P<0.05 was considered statistically significant. The following R packages facilitated building the different models:

(I) Multivariate adaptive regression splines: earth function in the earth package by Stephen Milborrow [2019] (12).

(II) Logistic regression model: inbuilt R function glm with logit family option.

Results

Characteristics of the total population

A total of 6,462 patients with T2DM were recruited based on the inclusion and exclusion criteria. The present study included 3,405 male subjects and 3,057 female subjects. The study participants ranged in age from 22 to 90 years, and the median was 57 [49–63] years. The prevalence of HbA1c improvement was 38.35% (2,478 patients). The prevalence of this improvement was higher for females than for males (39.19% *vs.* 37.59%). Lower BMI, less family history of T2DM, shorter duration, fewer related complications, fewer kinds of oral medicine, more exercise, adjusted diet habits, lack of hypertension, lower insulin dose, and lower 2hPBG were conducive to improving HbA1c (P<0.05). The characteristics of the variables in the two groups are listed in *Table 1*.

Multivariate adaptive regression splines

To explore the predictors for HbA1c improvement, a MARS model was employed using the above variables. The different basis functions and their coefficients are listed in *Table 2*. An insulin dose less than 13 U, more than 3 kinds of oral medicine, exercise frequency more than 1 time per week and 2hPBG less than 10.56 mmol/L were beneficial to improving HbA1c. By contrast, the following interactions were negatively associated with HbA1c improvement: having relative complications and 2hPBG less than 10.56 mmol/L, T2DM duration more than 7 years and insulin dose less than 13 U.

Once the MARS model was constructed, it was possible to evaluate the importance of the explanatory variables used to construct the basic functions. According to the

Table 1 Characteristics of the total population stratified by status of HbA1c improvement [n (%) or mean ± standard deviation]

Variables	Improvement group	Non-improvement group	Statistics	P
Gender			1.74	0.1874
Male	1,280 (37.59)	2,125 (62.41)		
Female	1,198 (39.19)	1,859 (60.81)		
Age (years)	54.82±11.23	56.88±11.16	7.19	<0.0001
BMI (kg/m²)	25.18±2.88	25.51±3.06	4.32	<0.0001
History of family of T2DM			10.32	0.0013
Yes	806 (35.70)	1,452 (64.30)		
No	1,672 (39.77)	2,532 (60.23)		
T2DM duration (years)	5.79±4.56	7.15±5.06	10.95	<0.0001
Relative complications, n (%)			111.41	<0.0001
Yes	972 (31.64)	2,100 (68.36)		
No	1,506 (44.42)	1,884 (55.58)		
Oral medicine (kinds)			74.93	<0.0001
0	223 (49.34)	229 (50.66)		
1	1,112 (42.28)	1,518 (57.72)		
2	974 (34.48)	1,851 (65.52)		
3	162 (30.06)	377 (69.94)		
4	7 (46.67)	8 (53.33)		
5	0 (0)	1 (100)		
Treat as directed			8.93	0.0028
Yes	2,365 (38.79)	3,732 (61.21)		
No	113 (30.96)	252 (69.04)		
Exercise frequency			126.92	<0.0001
0	314 (25.28)	928 (74.72)		
1 (≤2 times per week)	924 (38.71)	1,463 (61.29)		
2 (3–4 times per week)	988 (44.42)	1,236 (55.58)		
3 (≥5 times per week)	252 (41.38)	357 (58.62)		
Adjust diet habits			54.29	<0.0001
Yes	2,280 (39.97)	3,424 (60.03)		
No	198 (26.12)	560 (73.88)		
Hypertension			84.53	<0.0001
Yes	532 (29.42)	1,276 (70.58)		
No	1,946 (41.81)	2,708 (58.19)		
Insulin dose (U)	15.43±5.46	17.22±5.94	12.16	<0.0001
2hPBG (mmol/L)	8.57±1.86	11.37±2.79	44.23	<0.0001

1300

Lu et al. The application of MARS model in studying HbA1c improvement

Table 2 List of basic functions of the multivariate adaptive regression splines (MARS model) and their coefficients

Basic functions	Coefficients
(Intercept)	−1.785451
h(13-Insulin dose)	0.108270
h(Insulin dose-13)	−0.049137
h(3-Oral medicine)	0.430189
h(Oral medicine-3)	2.097661
h(1-Exercise frequency)	−0.384976
h(Exercise frequency-1)	0.040499
h(10.56-2hPBG)	1.012657
h(2hPBG-10.56)	−0.294497
Related complications * h(10.56-2hPBG)	−0.211556
h(49-Age) * h(10.56-2hPBG)	−0.016500
h(Age-49) * h(10.56-2hPBG)	−0.002273
h(T2DM duration-7) * h(13- Insulin dose)	−0.022545
h(0.8-T2DM duration) * h(10.56-2hPBG)	1.020654
h(T2DM duration-0.8) * h(10.56-2hPBG)	−0.008617
h(13-Insulin dose) * h(2hPBG-7.7)	0.024570
h(13-Insulin dose) * h(7.7-2hPBG)	0.010797
h(Insulin dose-13) * h(2hPBG-8.9)	0.009962
h(Insulin dose-13) * h(8.9-2hPBG)	0.039032
h(2-Oral medicine) * h(10.56-2hPBG)	−0.068879
h(Oral medicine-2) * h(10.56-2hPBG)	−0.134964
h(3-Oral medicine) * h(2hPBG-6.5)	−0.025979
h(3-Oral medicine) * h(6.5-2hPBG)	−0.786788

impact of HbA1c compliance in patients with T2DM, the importance of the variables was ranked as follows: 2hPBG, oral medicine, related complications, insulin dose, exercise frequency, T2DM duration and age (*Figure 1*).
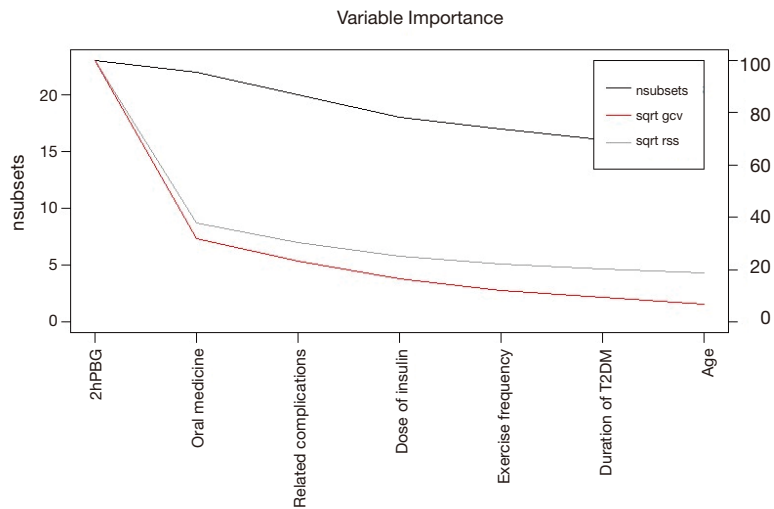
Model performance

The evaluation index of performance for the two models in the training and test sets is reported in *Table 3*. In the training test, the accuracy, Youden's index, recall rate, G-mean and AUC for the MARS model were all higher than those for logistic regression. The AUC for the MARS model was 0.843 (0.832–0.855), while the AUC for logistic regression was 0.836 (0.824–0.848), and the difference

in AUC was statistically significant. In the test set, the accuracy of logistic regression was slightly higher than that of MARS. The Youden's index, recall rate, G-mean and AUC of the MARS model were still higher, and the difference in AUC was also statistically significant. In a comparison of different data sets, the AUC for the MARS model decreased from 0.843 in the training set to 0.830 in the test set—a decrease of 0.013—and the decrease in AUC for the logistic regression was 0.129.

Discussion

HbA1c improvement was related to improved long-term outcomes and reduced diabetic complications. This study

Figure 1 The importance of explanatory variables in the multivariate adaptive regression splines model.

Table 3 Comparison of predictive performance of the multivariate adaptive regression splines (MARS) and logistic regression models

The evaluation index	Training set (n=4,524)		Test set (n=1,938)	
	MARS	Logistic regression	MARS	Logistic regression
Accuracy	0.773 (0.760–0.785)	0.760 (0.747–0.772)	0.762 (0.743–0.781)	0.763 (0.744–0.782)
Youden's index	0.528 (0.513–0.543)	0.430 (0.416–0.444)	0.494 (0.472–0.516)	0.413 (0.391–0.435)
Recall rate	0.850 (0.840–0.860)	0.747 (0.734–0.760)	0.850 (0.834–0.866)	0.756 (0.737–0.775)
G-mean	0.790 (0.778–0.802)	0.620 (0.606–0.634)	0.774 (0.755–0.793)	0.609 (0.587–0.631)
AUC	0.843 (0.832–0.855)[a]	0.836 (0.824–0.848)[b]	0.830 (0.812–0.849)[a]	0.707 (0.687–0.726)[b]

a, b: the different letters indicate a statistical significance between the two groups.

revealed that the prevalence of HbA1c improvement in five cities in northern China was 38.35%, which was lower than that in Guangdong and Shanghai, possibly because of the region's relatively lower economic level, low degree of medical insurance coverage and patients' self-management (13,14). This study also found that all variables other than gender showed statistically significant differences between the two groups (P<0.05). Therefore, reasonable control of the above factors can not only improve HbA1c but also have important clinical value in disease control and prevention for patients with diabetes.

The MARS model results show that 2hPBG, insulin dose, oral medicine, exercise frequency and T2DM duration all affected HbA1c improvement, which was consistent with the results of other relevant studies (5,15,16).

Many clinical studies have illustrated a positive correlation between 2hPBG and HbA1c (17). Theoretically, it can be explained that HbA1c is the product of an irreversible non-enzyme protein glycosylation reaction between glucose and hemoglobin in erythrocytes. This study found consistent results that 2hPBG <10.56 mmol/L is more conducive to improving HbA1c.

Insulin therapy is considered to be the most effective treatment method to reduce blood glucose, and a clinical study has shown that early, intensive insulin therapy reduces glucotoxicity and protects β cells (18). However, there is a common clinical phenomenon in China in which early doses of insulin are insufficient or even the initial time of insulin use is delayed. Therefore, a patient may have a lower insulin dose because he or she is in the early stage of diabetes and has had the disease for only a short time, which is conducive to controlling HbA1c. However, the interaction results

showed that patients with longer T2DM duration and lower insulin doses had more difficulty reducing HbA1c. It was suggested that for longer courses of illness, the treatment may be insufficient and lead to poor HbA1c improvement.

A multicenter observational study in Taiwan (5) found that in the early stage of diagnosis, the treatment of oral hypoglycemic agents combined with basal insulin was more conducive to controlling HbA1c. Riddle *et al.* investigated the effect of glycemic control with basal insulin-added oral medicine in patients with T2DM, and their results showed that the mean HbA1C levels at the end of the studies were all below 7% (19). Similar to those findings, our study found that the achievement of HbA1C <7% was positively related to having more than 3 kinds of basal insulin-added oral medicine.

Exercise has numerous health benefits. Regular exercise is widely considered important and can prevent the progression of diabetes. Many metabolic diseases seem to be attenuated by regular physical exercise (20) because it can consume a large amount of energy. Therefore, exercise is an effective way to lower blood glucose levels. Based on the study results, we recommend that patients with HbA1c ≥7% perform regular exercise weekly.

There is increasing interest in using classifiers to predict improvements in HbA1c. The MARS model used in this study is a powerful approach that is rarely used in public health and allows us to model the interaction of explanatory variables. The interaction between the influencing factors found in this study has not been reported in other studies. The accuracy, Youden's index, recall rate and AUC were all indicators to evaluate the properties of the models, and they were related to each other but with different emphases. Relatively speaking, the ROC curve was much more stable and showed a tradeoff between sensitivity and specificity (21,22). The MARS model performed better than logistic regression in the above aspects, except for accuracy. The AUC of logistic regression had a greater decline than the MARS model. A reasonable explanation is the tendency of the more complex logistic regression model to be over-fit on the training set. By contrast, the MARS model was more robust for improving the predictive ability of HbA1c. This finding was likely attributable to the fact that MARS employed generalized cross-validation in selecting terms for inclusion in the regression model. Hence, the main advantage of MARS was that it effectively modeled nonlinear relationships and patterns that could not be done by other regression methods. MARS could be applied in various types of studies because of its flexibility (23).

Conclusions

The prevalence of HbA1c improvement was low in the present study. Duration, 2hPBG, insulin dose, oral medicine, exercise frequency and diabetes complications all affected HbA1c improvement. Given the interaction between factors affecting HbA1c improvement, medical staff should conduct comprehensive interventions to further reduce HbA1c in patients. In this study, the MARS model was superior to the traditional logistic regression in improving HbA1c. MARS had greater generalizability because it not only considered nonlinear relations in the process of model fitting but also adopted cross-validation. Nevertheless, more studies are needed to provide evidence for this result.

Acknowledgments

Footnote

Data Sharing Statement: Available at http://dx.doi. org/10.21037/apm-19-406

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi. org/10.21037/apm-19-406). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and approved by Medical Ethics Committee of Tianjin Medical University. All the subjects had signed the informed consent forms.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with

the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Chinese preventive medicine association. Guidelines for the prevention and treatment of type 2 diabetes in China (2017 edition). Chin J Pract Intern Med 2018;38:292-344. (In Chinese)

2. Deconinck E, Zhang MH, Petitet F, et al. Boosted regression trees, multivariate adaptive regression splines and their two-step combinations with multiple linear regression or partial least squares to predict blood–brain barrier passage: A case study. Analytica Chimica Acta 2008;609:13-23.

3. Austin PC. A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. Stat Med 2007;26:2937-57.

4. Song YS, Koo BK, Kim SW, et al. Improvement of Glycosylated Hemoglobin in Patients with Type 2 Diabetes Mellitus under Insulin Treatment by Reimbursement for Self-Monitoring of Blood Glucose. Diabetes Metab J 2018;42:28-42.

5. Lin SD, Tsai ST, Tu ST, et al. Glycosylated hemoglobin level and number of oral antidiabetic drugs predict whether or not glycemic target is achieved in insulin-requiring type 2 diabetes. Prim Care Diabetes 2015;9:135-41.

6. Lorenzo-Medina M, Uranga B, Rus A, et al. Sex and age affect agreement between fasting plasma glucose and glycosylated hemoglobin for diagnosis of dysglycemia. Endocrinol Diabetes Nutr 2017;64:345-54.

7. American Diabetes Association, Standards of medical care in diabetes 2013. Diabetes Care 2013;36:S11-66.

8. Friedman JH. Multivariate adaptive regression splines. Ann Stat 1991;19:1-67.

9. Deconinck E, Xu QS, Put R, et al. Prediction of gastro-intestinal absorption using multivariate adaptive regression splines. J Pharm Biomed Anal 2005;39:1021-30.

10. Xu QS, Liang YZ. Monte Carlo Cross Validation. Chemom Intell Lab Syst 2001;56:1-11.

11. Friedman JH, Roosen CB. An introduction to multivariate adaptive regression splines. Stat Methods Med Res 1995;4:197-217.

12. Stephen Milborrow, 2019. Package 'earth'. Multivariate Adaptive Regression Splines. CRAN-R. Available online: http://www.milbo.users.sonic.net/earth

13. Song XH. Investigation on the Glycosylated Hemoglobin Control Effect and Influence Factors of 200 Patients with Diabetes Mellitus. Med innovation Chin 2017;14:58-61.

14. Jin LY, Wu G, Gong H. The influencing factors analysis of glycosylated hemoglobin among type 2 diabetic community patients. Shanxi J Med 2016;45:757-9.

15. Owen V, Seetho I, Idris I. Predictors of responders to insulin therapy at 1 year among adults with type 2 diabetes. Diabetes Obes Metab 2010;12:865-70.

16. Giugliano D, Maiorino M, Bellastella G, et al. Relationship of baseline HbA1c, HbA1c change and HbA1c target of <7% with insulin analogues in type 2 diabetes: a meta-analysis of randomised controlled trials. Int J Clin Pract 2011; 65:602–612.

17. Dubey D, Kunwar S, Gupta U. Mid-trimester glycosylated hemoglobin levels (HbA1c) and its correlation with oral glucose tolerance test (World Health Organization 1999). J Obstet Gynaecol Res 2019;45:817-23.

18. Retnakaran R, Drucker DJ. Intensive insulin therapy in newly diagnosed type 2 diabetes. Lancet 2008;371:1725-6.

19. Riddle MC, Rosenstock J, Gerich J. The treat-to-target trial: randomized addition of glargine or human NPH insulin to oral therapy of type 2 diabetic patients. Diabetes Care 2003;26:3080-6.

20. Nakhanakhup C, Moungmee P, Appell HJ, et al. Regular physical exercise in patients with type II diabetes mellitus. Eur Rev Aging Phys A 2006;3:10-9.

21. Hajian-Tilaki KO, Gholizadehpasha AR, Bozogzadeh S, et al. Body mass index and waist circumference are predictor biomarkers of breast cancer risk in Iranian women. Med Oncol 2011;28:1296-301.

22. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.

23. Lin HY, Wang W, Liu YH, et al. Comparison of multivariate adaptive regression splines and logistic regression in detecting SNP–SNP interactions and their application in prostate cancer. J Hum Genet 2008;53:802-11.