# Determining the factors related to diabetes type II with mixed logistic regression

Fariba Asadi[1*], Hossein Fallahzadeh[2], Masoud Rahmanian[3]

[1]Shahid sadoughi University of Medical Sciences, Yazd, I.R. Iran; [2]Biostatistics Dept., Shahid Sadoughi University of Medical Sciences, Yazd, I.R. Iran; [3]Diabetes Research Center, Shahid Sadoughi University of Medical Sciences, Yazd, I.R. Iran.

## ABSTRACT

**Background and aims:** Diabetes type II (non-insulin dependent) which is one of the most prevalent diabetes types in the world emerges in people with the age of above 55 and genetic and environmental factors interfere in this disease. The aim of this study was to determine the factors affecting diabetes type II with generalized mixed linear model.

**Methods:** Population of this study included 2820 people with the age of above 30 residing in Yazd Province who were selected using cluster sampling. To analyze the data, mixed logistic regression model was used in R software.

**Results:** In this study, 25% of men and 24.3% of women had diabetes. The regression analysis showed that age, WHR, family diabetes record, and BMI of 001 were the factors affecting diabetes, while variables of gender, house area, and education were not significant. On the other hand, unknown factors of residence place had high correlation with affliction with diabetes.

**Conclusion:** Based on the results obtained from this study, change of lifestyle and prevention of obesity can prevent affliction with diabetes to a great extent.

**Keywords:** Diabetes, Logistic regression, Mixed models, GLMM.

## INTRODUCTION

Diabetes which is one of the most prevalent chronic diseases has been increasing in recent years due to industrialization and change in the lifestyle of people and about 8% of Iranians suffer from it. Diabetes is induced by the shortage of insulin or reduction of its effect.[1] This disease is divided into two types: Diabetes I(insulin dependent) and diabetes II (non-insulin dependent). The latter is one of the most prevalent types of diabetes in the world and about 90-95% of people suffer from it.[2,3] This type of diabetes emerges almost in people with the age of above 55 and genetic and environmental factors are effective for it as well. Diabetes which is the fifth cause of death in Western societies has been reported to have the prevalence of 1-4% and include 10% of total medical emergency care.[4] World Health Organization has announced diabetes as a hidden epidemic disease and invited all

*Corresponding author: Fariba Asadi. Student, Shahid sadoughi University of Medical Sciences, Yazd, I.R. Iran, Tel: 00989355435603, E-mail: fariba.asadi3@gmail.com

countries of the world to fight against it since 1993. In a study on health and disease which was conducted in 1999, its prevalence was reported approximately 1.5 in Iran. At present, the number of people with this disease in the world is estimated about 285 million people. If a suitable strategy is not undertaken for its prevention, this number will increase from 135 million people in 1995 to 335 million people in 2025 according to WHO estimate.[4]

Many patients with diabetes type II do not show symptoms of this disease and are not informed of their disease for years. Studies have shown that, when this disease is identified in a person, he has already suffered from it for 4 to 7 years. When this disease is diagnosed in the patients with diabetes II, 25, 9, and 8% will suffer from retinopathy, neuropathy, and renal diseases, respectively.[5] Therefore, this disease is a challenge in the health system and its effective factors and level of awareness among people in the society should be identified. Sometimes due to data sampling and collection way such as cluster sampling or longitudinal studies, there is a type of data correlation. In other words, it can be said that findings about people inside a cluster or findings of about person at different times were more similar to those about the people inside another cluster than the findings about another person. The lack of consideration of this correlation may be caused usually underestimated and inaccurate results. Independence of data and normality hypothesis are the infrastructures of most common statistical methods.[6] Considering these hypotheses, many problems can be analyzed; however, there are many situations which violate main hypotheses of independence and normality of data.

In the longitudinal studies which are abundant in medicine, epidemiology, economy, agriculture, and other fields, the expectation that there should be a correlation between findings of the response variable of each studied sampling unit over time leads statisticians to find models which model this correlation. For this reason, normal linear mixed model was introduced in 1982 which has been widely applied in the analysis of longitudinal data.[7] In this model, considering a random variable for each sampling unit enables the explanation of the correlation inside each unit (among the findings over time).

In fixed time effect model, it is assumed that all findings are independent from each other, which leads to unsuitable models for the analysis of all types of correlated data, particularly longitudinal and clustering data. Generalized linear mixed models are a combination of generalized linear models[8] and linear mixed models[6,8-12], which include the added components from variability due to the presence of hidden random effects.[13-15] The reason that cluster is not considered in these models as fixed effect is that there are many clusters in the studies which cause inclusion of many parameters, and high complexity of the model, and incorrect results. If cluster is considered to have random effects, then only one conditional parameter will describe dispersion in the model.[11] General form of these models is, where X is a vector of fixed effects and Z is a vector of random effects. In these models, a distribution is determined for Y|b, instead of specifying a distribution for Y; i.e. response variable as a state of the generalized linear models.[16] The aim of this study was to identify the factors affecting

330

diabetes in city of Yazd. Since sampling method in this study was Cluster sampling, classic linear models cannot be applied for its analysis. As a result, it seems necessary to use generalized mixed linear model. Also, this data has already not been analyzed with an accurate method of statistical, using mixed logistic regression. It will be possible to compare this study with cluster sampling studies that conducted carefully in future.

## METHOD

In this study, data relating to the study of epidemiological indices of adult diabetes in the age group of 30 years and above in Yazd Province in 1998 were used. In the conducted study, the samples were selected using clustering method and total number of the samples was 2795 people aging more than 30 years old who were studied in 947 clusters, each with 20 families. After the selection of the clusters, the selected houses were visited and the initial questionnaire was completed for the people of the family aging above 30 years old. On the next morning, 12 h fasting blood of the sample was taken and then 75 g of glucose was solved in 300 cc of water and was given to them within 3-5 min.The second blood sample was taken for 2-h sugar after 2 h. After the centrifuge and isolation of serum, blood sugar was measured enzymatically. In this study, the people who had fasting blood sugar of 110-140 mg/dl or 2-h blood sugar of 140 to 200 mg/dl after the intake of 75 g glucose were invited to perform Oral Glucose Tolerance Test (OGTT). OGTT test was performed for 136 out of 301 people who were at the first IGT stage. Diabetes and IGT

were divided based on the criteria of WHOM. Since data was collected using clustering sampling method, then it was necessary to use the mixed model which is very applicable and useful in the analysis of clusters and longitudinal data. In this study, variables of gender, BMI, WHR, education level in two lower high school, high school, and upper high school degrees, job, area of house, family diabetes record, and age (including four classesof above 65, 50-64, 40-49, and 30-39) were studied as independent variables and affliction with diabetes (yes or no) was studied as binary response variable. Random effects including area and family were studied. Then, the mixed logistic regression model was fitted into the data in R software with lme4 package using Laplace method.

## RESULTS

Table 1 shows diabetes condition of the residents in Yazd Province in terms of gender and place of residence. On this basis, 25% of the people had diabetes, which was more prevalent among men than women; but this value was not statistically significant (P>0.1). As observed in the table, prevalence of diabetes was different in different cities so that Bafgh had minimum and AbarKooh had maximum prevalence (34.7% versus 6.7%). This difference was statistically significant (P<0.001). Table 2 indicates diabetes in terms of BMI and age groups. Considering the results in this table, it can be understood that, with increasing age and also index of BMI, the percentage of the diabetic people increased and this difference was statistically significant (P<0.01).

**Table 1:** Diabetes condition in terms of gender and place of residence

| Variable | | Affliction with diabetes | |
|---|---|---|---|
| | | Yes | No |
| Gender | Female | 24.7% | 75.3% |
| | Male | 25.4% | 74.6% |
| Area of house | Yazd | 25.8% | 74.2% |
| | Taft | 24.2% | 75.8% |
| | Meybod | 16.2% | 83.8% |
| | Ardakan | 29.9% | 70.1% |
| | Mehriz | 25.5% | 74.5% |
| | Bafgh | 34.7% | 65.3% |
| | Abarkooh | 6.7% | 93.3% |
| Total | | 25% | 75% |

**Table 2:** Percentage of people with diabetes in terms of variables of age and BMI

| Variable | BMI | | | | Age groups | | | |
|---|---|---|---|---|---|---|---|---|
| | Thin | Normal | Overweigh | Obese | 30-39 | 40-49 | 50-64 | 65+ |
| Percent of people with diabetes | 18% | 19.8% | 69.3% | 27.8% | 12.1% | 21.2% | 37.4% | 75.4% |

Results obtained from fitting of mixed logistic regression model which is given in Table 3 indicated very high prevalence of diabetes in the people with higher BMI than normal and thin people. Considering this table, variables of age, family record, BMI, and WHR became significant at $P<0.001$ level. In other words, the factors relating to diabetes were diagnosed while the factors of house area, gender, job, and education did not become significant. On the other hand, the unknown factors relating to place of residence and the unknown factors relating to any family covered 38 and 12% of changes in affliction with diabetes.

Among the random effects only significant effect was the location ($P<0.001$). Considering the table, the high odds ratio relating to WHR and is equal to 6.58 which indicated 6.58% increase of odds ratio affliction with diabetes with one unit of increase in this index was observed.

**Table 3:** Odds ratio of affliction with diabetes based on fitting of mixed logistic regression model

| Variable | | Estimate value | Odds ratio | P |
|---|---|---|---|---|
| Family history | | -0.70 | 0.49 | <0.001 |
| Age groups[1] | 40-49 | 0.55 | 1.73 | <0.001 |
| | 50-64 | 1.41 | 4.12 | <0.001 |
| | 65+ | 1.87 | 6.53 | <0.001 |
| BMI[2] | Normal | 0.32 | 1.38 | 0.12 |
| | Overweight | 0.92 | 2.25 | <0.001 |
| | Obese | 0.83 | 2.31 | <0.001 |
| Education level[3] | | -0.15 | 0.85 | 0.31 |
| WHR | | 1.88 | 6.58 | <0.001 |
| Area of house | | 0.00 | 1.00 | 0.18 |
| Gender[4] | | 0.00 | 0.99 | 0.97 |
| Job[5] | Worker | -0.23 | 0.79 | 0.33 |
| | Homemaker | -0.09 | 0.90 | 0.61 |
| | Famer | -0.47 | 0.62 | 0.18 |
| | Pensionary | -0.28 | 0.74 | 0.24 |
| | Other | -0.06 | 0.93 | 0.73 |
| Variance of first-level effect | | 0.12 | | 0.4 |
| Variance of two-level effect | | 0.38 | | <0.001 |

*Reference groups,respectively: 1.(30-39) 2.thin 3.under of diploma 4.male 5.emploee.*

## DISCUSSION

Considering the results obtained from this paper, it can be observed that variables of age, family record, BMI, and WHR were the factors affecting affliction with diabetes. Unknown factors of residence place had high correlation with affliction with diabetes. Kelestimur et al studied the prevalence of diabetes and risk factors related to diabetes type II in people aging above 30 years old in Kayser city of Turkey. Results showed that family record, hypertension, and obesity were significant in affliction with diabetes.[17] GholamrezaVaghari et al. in a descriptive study investigated the prevalence of diabetes type II and some related factors in adults aging between 25 and 65 years old in Golestan Province. In this study whose samples had been collected using clustering and stratified sampling method, Chi-squared test and t-test were used for the analysis of data. Results showed that the prevalence of diabetes increased with increasing age, reducing physical activity and abdominal obesity, and increasing BMI. No significant relationship was found between economic status and literacy level and affliction with diabetes.[18] As was mentioned, linear mixed models prevent mistaken significance of many independent variables considering the correlation between the clusters in case this correlation is high and

provides more accurate results for researchers. In a case-control study was conducted in Kurdistan to analyze the data was used, chi-square test, Mantel-Hansel test. The results showed that obesity hypertension, age ≥ 40, sex and history of stillbirths were related to diabetes. Also, Mohammad Reza Merati et al. conducted a cross-sectional study to investigate age and gender prevalence of hypertension and diabetes and studied some of their risk factors on 3000 men and women aged from 15 to 64 years old. Results showed that prevalence of diabetes in women was higher than that of men and there was a significant relationship between age, gender, hypertension, BMI, and family record on the one hand and affliction with diabetes on the other.[19] This study was conducted using stratified sampling method along with multistage clustering. Multivariate logistic regression was used for the analysis of data. Probably, it can be said that the reason for the significance of sex in these two studies and its difference from our study can be attributed to the lack of consideration of correlation between clusters. Skronal et al. in a study, concluded that the lack of correlation in the data would underestimate the standard error of the regression coefficients.[20] Also, in this study, the family random effect was not significant but location random effect was highly significant. This result reflects the high correlation between Location unknown factors of people with diabetes. Although many studies have been conducted on the factors affecting diabetes in Yazd Province, results of this study were more significant and generalizable and none of them were more comprehensive than this study.

# CONCLUSION

Based on the results obtained from this study, change of lifestyle and prevention of obesity can prevent affliction with diabetes to a great extent.

# CONFLICT OF INTEREST

The authors declare no conflict of interest.

# ACKNOWLEDGEMENTS

# REFERENCES

1. Levitt NS, Steyn K, Lambert EV, Reagon G, Lombard CJ, Fourie JM, et al. Modifiable risk factors for Type 2 diabetes mellitus in a peri-urban community in South Africa. Diabet Med. 1999; 16(11): 946-50.

2. Harris M. Definition and classification of diabetes mellitus and the new criteria for diagnosis. Diabetes Care. 2010; 33(Suppl 1): S62-S69.

3. Bergenstal R, Kendall D, Franz M, Rubenstein A. Management of type 2 diabetes: A systematic approach to meeting the standards of care Self-management education medical nutrition therapy and exercise. Endocrinology 4th edition Philadelphia: WB Saunders Company. 2000: 810-20.

4.Nakhodayi zade M, Raissi dehkardi F, Babolkhani E. Comparison healthy lifestyle compared to type 2 diabetes in Shohada hospital in the city of Khorramabad in 1387.

2nd International Congress of Metabolic Syndrom Obesety and Diabetes; 2010.

5.What is diabetes? Comprehensive base of Medical Information of Iran; 2014.

6. CHarles E, McCulloch. An introduction to generalized linear mixed models. Annual Conference on Applied Statistics in Agriculture, Biometrics Unit and Statistc Center; 1997.

7. Bates D. Fitting linear mixed models in R. R news. 2005; 5(1): 27-30.

8. Venables WN, Dichmont CM. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fish Res. 2004; 70(2): 319-37.

9. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MH, et al. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol. 2009; 24(3): 127-35.

10. Booth JG, Hobert JP. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. J R Stat Soc Series B. 1999; 61(1): 265-85.

11. Agresti A, Kateri M. Categorical data analysis: Springer; 2011.

12. Jiang J. Linear and generalized linear mixed models and their applications: Springer Science and Business Media; 2007.

13. Zhu H-T, Lee S-Y. Analysis of generalized linear models via a stochastic approximation algorithm with Markov chain Monte-Carlo method. Comput Stat. 2002; 12(2): 175-83.

14. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. J Am Stat Assoc. 1993; 88(421): 9-25.

15. Gilani N, Kazemnejad A, Zayeri F, Yazdani J. Comparison of Marginal Logistic Model with Repeated Measures and Conditional Logistic Model in Risk Factors Affecting Hypertension. J Mazandaran Univ Med Sci. 2011; 21(82): 27-35.

16. Fitzmaurice GM, Laird NM, Ware JH. Generalized Linear Mixed Models. applied longitudinal analysis: USA: John Wiley and Sons; 2012.

17. Kelestimur F, Cetin M, Pasaoglu H, Coksevim B, Cetinkaya F, Unluhizarci K, et al. The prevalence and identification of risk factors for type 2 diabetes mellitus and impaired glucose tolerance in Kayseri, central Anatolia, Turkey. Acta Diabetol. 1999; 36(1-2): 85-91.

18. Vaghari G, Sedaghat S, Joshaghani H, Hosseini SA, Niknejhad F, Angize A, et al. The prevalence of type II diabetes and associated risk factors in adults aged 25 to 65 years in Golestan Province. J Res Nurs Midwifery. 2011; 7(1): 69-74.

19. Merati M, Feizei A, Bager Nejad M. Prevalence of high blood pressure and diabetes and risk factors associated with them, based on a large study of the general population- an application of multivariate logistic regression models. Health Syst Res. 2012; 8(2): 193-203.

20. Skronal DHS. Multilevel logistic regression. Stat Med. 2002; 3: 411-20.