Original Research Article

Exploring the Risk Factors of Diabetes in Dhaka City: Negative Binomial Regression & Logistic Regression Approach

Mohammad Ahsan Uddin[*]

Associate Professor, Department of Statistics, University of Dhaka, Dhaka, Bangladesh

Abstract

In the developing countries, diabetes has become one of the most indicating public health challenges. It imposes a big economic impact on the society. The Bangladeshi people hardly understand the risk factors that causes diabetes. This study mainly aims to explore the potential risk factors behind this disease through some popular statistical analysis approaches. In this regard, a field survey was conducted on the diabetes patients in Dhaka city. The analysis involves binary logistic regression, negative binomial regression, and multinomial logistic regression model according to the nature of the data. The variables family history, height, weight, gender, age, food with high cholesterol, no exercise, taking alcohol/cigarettes, hypertension, and eye problem are found significant in this study. In addition, the count data related to the number of time(s) a patient checks blood glucose level is also analyzed and interpreted. From the overall findings, the study suggests to become extra careful of the disease who have already diabetes patient(s) in the family. Increased height and losing weight have good advantages to reduce the risk of diabetes. Avoiding high cholesterol food and cigarettes/alcohol is a wise decision in this regard. The lower risk of diabetes depends on regular exercising and controlling hypertension. The adults and the obese people are more serious of checking their blood glucose level. Educated persons are also careful of this matter. Since most of the risk factors for diabetes are preventable, efforts should be taken by the Government to better understand of these factors to the public.

Keywords: diabetes, risk factors, statistical analysis, binary logistic regression, negative binomial regression, multinomial logistic regression, count data.

INTRODUCTION

Recently, the burden of diabetes is increasing rapidly. Diabetes is identified as one of the four major types of non-communicable diseases (NCDs) which have a large contribution to worldwide morbidity and mortality [1] . Several complications like diabetic neuropathy, retinopathy, diabetic foot ulcer, nephropathy, etc. are being emerged from this disease causing serious harm to public health[2]. It also leads to chronic kidney diseases, stroke, neuropathy, heart attack, visual impairment, etc. and thus becomes responsible for the decreased quality of life creating a serious long-term effect on financial and social wellbeing of the affected individuals[3].

Bangladesh has been identified as the top 10 high-burden diabetes countries in the world having an estimated 8.4 million people with diabetes and another 7.8 million with pre-diabetes [4] . In addition, it is projected that Bangladesh will rank 5th in the world for the highest growth in diabetes population by 2030[5].

This high incidence of diabetes is supposed to produce demolishing social and economic consequences on the healthcare systems of the country[6]. In a developing country like Bangladesh, the financial impacts that emerge from diabetes to the individuals and families are much severe which in terms affect the whole economy causing loss of productivity, mortality, and morbidity[6 − 7].

The first objective of the present study is to determine the causes of diabetes by analyzing a sample of diabetes patients of Bangladesh from Dhaka city. Binary logistic regression model, a very common statistical analysis approach, is used in this purpose. The number of blood glucose level checking by the diabetes patients is analyzed with negative binomial regression analysis approach and the significant responsible factors that determine this counted number is sought. Moreover, logistic regression model is considered in the last part of the study to find out the determiners of having diabetes compared to occurring the disease by inherited. Also, the recommendations

made by the study are discussed in the concluding remark.

METHODS AND MATERIALS

A primary dataset has been collected from Dhaka city for the study purpose. Purposive sampling was involved in this regard and a questionnaire, covering the study objectives, was made accordingly. 16 undergraduate students worked in teams as a voluntary service to collect the data. Direct interview (face-to-face) method was used while collecting data and respondents were selected according to the ease of the interviewers. Data collection continued from 10 August 2019 to 12 September 2019 and covered the required responses from 900 respondents, 450 having diabetes and rest are free from the disease.

Three different forms of regression analysis are considered in this study for analysis purpose. Regression analysis is a popular statistical tool to investigate the relationships between a dependent variable and a set of covariates[8]. It determines which factors are important, which factors can be ignored, and how these factors influence each other[9]. The analysis is mostly driven by three important facts (number of independent variables, type of dependent variables, and shape of regression line[10]. If the response vector $\mathbf{Y}$ be of binary type i.e., referring to whether an event has occurred or not, binary logistic regression is used for modeling purpose, which has the following form [11]

$$\pi(\mathbf{x}) = \frac{e^{\mathbf{x}'\boldsymbol{\beta}}}{1 + e^{\mathbf{x}'\boldsymbol{\beta}}}$$

Where $\pi(\mathbf{x})$ represents the conditional mean of $\mathbf{Y}$ given $\mathbf{x}$ i.e., $E(\mathbf{Y}|\mathbf{x})$. The unknown parameters $(\boldsymbol{\beta})$ are estimated by the method of maximum likelihood estimation[12].

Another form of regression analysis is negative binomial regression model which is used to model count data [13]. To fit this regression model, the response variable has to follow a negative binomial distribution and the logarithm of its expected value needs to be expressed as a linear combination of unknown parameters[14]. Mathematically, if Y has a negative binomial distribution then [15]

$$\log\big(E(Y|\mathbf{x})\big) = \alpha + \boldsymbol{\beta}'\mathbf{x},$$

Where α is the intercept term and $\boldsymbol{\beta}$ be the vector of regression parameters corresponding to the vector of independent variables $\mathbf{x}$. The unknown parameters $\boldsymbol{\beta}$ can be estimated by maximum likelihood estimation when the response values are distributed independently[16].

Multinomial logistic regression is another common statistical tool to predict the outcome of a categorical dependent variable based on the predictor variable(s) [17]. For qualitative response variable, multinomial logistic regression is used to estimate the empirical values of the unknown parameters[18]. The multinomial logistic regression model can be expressed as

$$E(Y = 1 \mid \mathbf{x}) = \frac{1}{1 + e^{-(\alpha + \boldsymbol{\beta}'\mathbf{x})}},$$

Where the unknown parameters (as stated in binomial logistic model) are estimated by maximum likelihood approach[19].

RESULTS AND DISCUSSIONS

The potential determinants that are responsible for occurring diabetes are shown in Table 1. The variables family history, weight, and age turned out as significant in the model, whereas, mental pressure and gender are insignificant. It is found from Table 1 that the variable family history is significant at 5% level of significance. An individual who has diabetes patient(s) in his family is 1.217 times as likely to have diabetes compared to an individual having no diabetes patient in the family, keeping all other variables at a fixed level. The analysis found that weight is significantly determining the occurrence of diabetes with p-value 0.031. One kg increase in weight results in 1.2% more odds of having diabetes. For one-year increase in age, the odds of having diabetes significantly decreases by 0.8% with p-value 0.034.

Table-1: Binary logistic regression model estimates of the selected covariates for diabetes disease in Dhaka city along with standard error (SE), odds ratio (OR), and p-value

Covariates	Coefficient	SE	OR	p-value
Constant	-.082	.579	0.921	.005
Family history				
Yes	.197	.220	1.217	.043
No	-	-	-	-
Mental pressure				
Yes	-.329	.237	0.719	.166
No	.239	.252	1.269	.344
Sometimes	-	-	-	-
Weight	.012	.007	1.012	.031
Age	-.008	.006	0.992	.034
Gender				
Male	-.042	.146	0.958	.772
Female	-	-	-	-

Table 2 is representing the factors that determine the number of blood glucose level checking. The variables education, occupation, age, and weight are found significant in the model. The insignificant variables in the model are exercise, medicine, and blood sugar. For the variable education, it is found from the analysis that the log of mean number of times blood glucose checking for an illiterate respondent is 0.309 unit lower than that of a post graduate respondent at 5% significance level (p-value 0.023), when all other variables are at a fixed level. In the same way, the respondents having primary education, secondary education, higher secondary education, and graduation have 0.232, 0.256, 0.142, and 0.172 unit, respectively, significant lower log of mean number of times blood glucose checking compared to post graduate

respondents having respective p-values 0.039, 0.093, 0.053, and 0.085. The variable occupation is significant in the model and reveals that the log of mean number of times blood glucose checking for retired persons is 0.069 unit lower significantly than the housewives having p-value 0.014. The teachers and others have, respectively, 0.151 and 0.301 unit significantly higher log of mean number of times blood glucose checking compared to the housewives with respective p-values 0.002 and 0.084. Again, it is observed from the result that one-year increase in age results in increasing 0.005 unit log of mean number of times blood glucose checking significantly with p-value 0.012. Weight is also significant in the model having p-value 0.059. One kg increase in weight increases 0.003 unit significant log of mean number of times blood glucose checking.

Table-2: Negative binomial model estimates of the selected covariates for diabetes disease in Dhaka city along with standard error (SE) and p-value

Covariates	Coefficients	SE	p-value
Constant	1.953	.338	.000
Education			
Illiterate	-.309	.159	.023
Primary	-.232	.157	.039
Secondary	-.256	.152	.093
Higher secondary	-.142	.153	.053
Graduate	-.172	.161	.085
Post graduate	-	-	-
Occupation			
Service holder	-.066	.156	.673
Retired	-.069	.149	.014
Businessman	.089	.149	.553
Teacher	.151	.173	.002
Others	.301	.174	.084
Housewife	-	-	-
Exercise			
Yes	-.118	.079	.134
No	-	-	-
Medicine			
Insulin	-.046	.119	.695
Pill	-.062	.118	.599
Both	-	-	-
Age	.005	.003	.012
Weight	.003	.004	.059
Blood sugar	-.003	.007	.696

Table 3 is depicting the results obtained from multinomial logistic regression model and the significant variables are interpreted here. In this model, the dependent variables are food with high cholesterol, no exercise, taking cigarettes/alcohol, hypertension, and cause of diabetes by inherited. The last one is considered as a reference dependent variable with which the rests are compared. At first, food with high cholesterol for causing diabetes compared to cause of diabetes by inherited is discussed and found height, weight, and gender as significant in model. The variable height is significant in the multinomial logistic estimate with p-value 0.052. For one cm increase in height, the odds of having diabetes of food with high cholesterol compared to by inherited decreases (1-

OR)*100%=3.1% significantly. One kg increase in weight significantly increases the log odds of having diabetes of food with high cholesterol compared to by inherited by 0.039 unit with p-value less than 0.001. Males are 92.2% significantly more likely than females to develop diabetes for food with cholesterol as a cause of diabetes compared to cause of diabetes by inherited having p-value 0.001. When no exercise is considered as a cause of diabetes compared to cause of diabetes by inherited, the variables age, height, and gender are obtained as significant. For one-year increase in age, the log odds of having diabetes from no exercise compared to by inherited increases by .013 units significantly at 10% significance level with p-value 0.096. No exercise is significantly 3.8% more likely to

occur diabetes compared to by inherit for one cm increase in height. The male respondents are at higher risk to develop diabetes for no exercise as the result shows that they have 28% significantly more odds of developing diabetes than the female respondents having p-value less than 0.001. The variables height, weight, and gender are obtained significant for causes of diabetes due to taking cigarettes/alcohol compared to cause of diabetes by inherited. For one cm increase in height the log odds of having diabetes for taking cigarettes/alcohol compared to by inherited decreases significantly by 0.06 unit with p-value 0.008. Respondents taking cigarettes/alcohol are significantly 3.5% less likely (p-value 0.070) for having diabetes other than by inherited if weight increases by one kg. Again, males are 0.2% significantly more likely than females to develop diabetes due to cigarettes/alcohol compared to cause of diabetes by inherited having p-value less than 0.001. When hypertension is considered as a cause of diabetes compared to by inherited, the variables height, gender, and eye problem become significant in the model. For one cm increase in height, the log odds of having diabetes due to hypertension compared to by inherited decreases significantly by 0.032 unit with p-value 0.097. Males are 74% significantly (p-value 0.031) more likely than females for developing diabetes due to hypertension compared to by inherited. The log odds for the individuals having eye problem is 0.877 times than the individuals having no eye problem for having diabetes due to hypertension compared to by inherited.

Table-3: Multinomial logistic regression estimates of the selected covariates for diabetes disease in Dhaka city along with standard error (SE), odds ratio (OR), and p-value

Causes of diabetes	Covariates	Coefficient	SE	OR	p-value
Food with high cholesterol	Intercept	-.526	1.080	0.590	.626
	Age	-.004	.007	0.996	.635
	Height	-.031	.016	0.969	.052
	Weight	.039	.010	1.039	<.001
	Gender				
	Male Male	.653	.199	1.921	.001
	Female	-	-		-
	Eye problem				
	Yes	.318	.202	1.374	.116
	No	-	-		-
	Intercept	.818	1.103	2.265	.458
	Age	.013	.008	1.013	.096
	Height	-.038	.016	0.962	.020
	Weight	.003	.010	1.003	.733
	Gender				
	Male Male	.247	.209	1.280	<.001
	Female	-	-		-
	Eye problem				
	Yes	.143	.211	1.153	.499
	No	-	-		-
Taking cigarettes/alcohol	Intercept	1.905	1.804	6.719	.291
	Age	.020	.016	1.020	.200
	Height	-.060	.023	0.941	.008
	Weight	-.035	.020	0.965	.070
	Gender				
	Male Male	.002	.441	1.002	<.001
	Female	-	-		-
	Eye problem				
	Yes	-.214	.405	0.807	.598
	No	-	-		-
Hypertension	Intercept	2.053	1.274	7.791	.107
	Age	-.008	.009	0.992	.405
	Height	-.032	.019	0.968	.097
	Weight	-.018	.012	0.982	.140
	Gender				
	Male Male	.554	.257	1.740	.031
	Female	-	-		-
	Eye problem				
	Yes	.877	.291	2.403	.003
	No	-	-		-

The determinants of diabetes in Dhaka city are measured in this study using rigorous statistical procedures. The result shows that having diabetes of a family member significantly increases the risk of developing this disease in other members of the family. This finding is very similar to the previous studies [19 − 20]. Increase in weight is significantly associated with the increasing chance of developing diabetes which supports a similar study on diabetes[21]. According to this study, the more you are getting older, the more you possess the risk of having diabetes[22]. From the analysis result of number of times blood glucose checking, it is obtained that the higher education one has, the more he/she checks blood glucose level [23]. The housewives have a lesser tendency to check blood glucose level than the teachers and other occupied patients though the retired persons have the lowest frequency of checking. But, increase in age increases the frequency of checking blood glucose level. The study also finds that the diabetes patients having higher body weight have more tendencies to check their glucose level.

CONCLUSION

The study reveals that taking high cholesterol food is significantly related to diabetes which is found similar to another study[24]. The respondents who take high cholesterol food have lower risk of diabetes with increase in height compared to cause of diabetes by inherited. But, in this situation, increasing weight is responsible for higher chance of developing diabetes. The analysis also gives evidence that taking high cholesterol food is more harmful to the males than the females to develop diabetes. Regular exercising is important to prevent diabetes [25] as it is found in this study that with the increase in age, the chance of having diabetes rises to the individuals taking no exercise compared to developing the disease by inherited. In this respect of taking no exercise, the male respondents have higher risk to diabetes and increasing height helps to decrease this risk. Taking cigarettes/alcohol is significantly associated with diabetes compared to the cause of diabetes by inherited[26]. Increase in height may decrease the chance of having diabetes of the cigarettes/alcohol taking respondents, while weight increasing is acting to decrease this chance compared to the cause of diabetes by inherited. In this case, the males are more likely to have diabetes than the females. Hypertension is also found significant in this study to occur diabetes [27]. For the dependent variable hypertension, the variables height and gender gives similar results as in the case of taking cigarettes/alcohol. Hypertension is significantly responsible to develop diabetes compared to by inherit to the respondents having eye problem. To sum up, the males are more prone to develop diabetes than the females and increased height is helpful to lessen the chance of having the disease.

Diabetes has been turned into a dangerous impendence to the health system of Bangladesh. This study seeks for the risk factors of diabetes based on the diabetes patients in Dhaka city. From the overall findings, the study suggests to become extra careful of the disease that has already diabetes patient(s) in the family. Increased height and losing weight have good advantages to reduce the risk of diabetes. Avoiding high cholesterol food and cigarettes/alcohol is a wise decision in this regard. The lower risk of diabetes depends on regular exercising and controlling hypertension. The adults and the obese people are more serious of checking their blood glucose level. Educated persons are also careful of this matter. Since most of the risk factors for diabetes are preventable, efforts should be taken by the Government to better understand of these factors to the public.

REFERENCES

1. Islam, S. M. S., Lechner, A., Ferrari, U., Laxy, M., Seissler, J., Brown, J., & Holle, R. (2017). Healthcare use and expenditure for diabetes in Bangladesh. BMJ global health, 2(1).
2. Islam, S. M. S., Niessen, L. W., Seissler, J., Ferrari, U., Biswas, T., Islam, A., & Lechner, A. (2015). Diabetes knowledge and glycemic control among patients with type 2 diabetes in Bangladesh. SpringerPlus, 4(1), 284.
3. Mohiuddin, A. K. (2019). Diabetes fact: Bangladesh perspective. International Journal of Diabetes Research, 2(1), 14-20.
4. Federation, I.D., Atlas, I. (2013). International Diabetes Federation. IDF diabetes atlas, 6th edn Brussels, Belgium: International Diabetes Federation.
5. Whiting, D. R., Guariguata, L., Weil, C., & Shaw, J. (2011). IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes research and clinical practice, 94(3), 311-321.
6. Islam, S. M. S., Lechner, A., Ferrari, U., Froeschl, G., Niessen, L. W., Seissler, J., & Alam, D. S. (2013). Social and economic impact of diabetics in Bangladesh: protocol for a case–control study. BMC Public Health, 13(1), 1-9.
7. Mørkrid, K., Ali, L., & Hussain, A. (2010). Risk factors and prevalence of diabetic peripheral neuropathy: A study of type 2 diabetic outpatients in Bangladesh. International journal of diabetes in developing countries, 30(1), 11.
8. Seber, G.A., Lee, A.J. (2012). Linear regression analysis: John Wiley & Sons.
9. Fox, J. (1997). Applied regression analysis, linear models, and related methods: Sage Publications, Inc.
10. Montgomery, D.C., Peck, E.A., Vining, G.G. (2012). Introduction to linear regression analysis: John Wiley & Sons.

11. Midi, H., Sarkar, S. K., & Rana, S. (2010). Collinearity diagnostics of binary logistic regression model. Journal of Interdisciplinary Mathematics, 13(3), 253-267.

12. McDonald, B. W. (1993). Estimating logistic regression parameters for bivariate binary data. Journal of the Royal Statistical Society: Series B (Methodological), 55(2), 391-397.

13. Hilbe, J.M. (2011). Negative binomial regression: Cambridge University Press.

14. Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychological bulletin, 118(3), 392.

15. Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. The Canadian Journal of Statistics/La Revue Canadienne de Statistique, 209-225.

16. Deb, P., & Trivedi, P. K. (2006). Maximum simulated likelihood estimation of a negative binomial regression model with multinomial endogenous treatment. The Stata Journal, 6(2), 246-255.

17. Böhning, D. (1992). Multinomial logistic regression algorithm. Annals of the institute of Statistical Mathematics, 44(1), 197-200.

18. Starkweather, J., & Moske, A. K. (2011). Multinomial logistic regression. Consulted page at September 10th: http://www. unt. edu/rss/class/Jon/Benchmarks/MLR_JDS_Aug201 1. pdf, 29, 2825-2830.

19. Harrison, T. A., Hindorff, L. A., Kim, H., Wines, R. C., Bowen, D. J., McGrath, B. B., & Edwards, K. L. (2003). Family history of diabetes as a potential public health tool. American journal of preventive medicine, 24(2), 152-159.

20. Loomba, R., Abraham, M., Unalp, A., Wilson, L., Lavine, J., Doo, E., ... & Nonalcoholic Steatohepatitis Clinical Research Network. (2012). Association between diabetes, family history of diabetes, and risk of nonalcoholic steatohepatitis and fibrosis. Hepatology, 56(3), 943-951.

21. O'Sullivan, J. B. (1982). Body weight and subsequent diabetes mellitus. Jama, 248(8), 949-952.

22. Sayeed, M. A., Hussain, M. Z., Banu, A., Rumi, M. A. K., & Khan, A. A. (1997). Prevalence of diabetes in a suburban population of Bangladesh. Diabetes research and clinical practice, 34(3), 149-155.

23. Terent, A., Hagfall, O., & Cederholm, U. (1985). The Effect of Education and Self- monitoring of Blood Glucose on Glycosylated Hemoglobin in Type I Diabetes: A Controlled 18- Month Trial in a Representative Population. Acta medica Scandinavica, 217(1), 47-53.

24. Tajima, R., Kodama, S., Hirata, M., Horikawa, C., Fujihara, K., Yachi, Y., ... & Sone, H. (2014). High cholesterol intake is associated with elevated risk of type 2 diabetes mellitus–A meta-analysis. Clinical nutrition, 33(6), 946-950.

25. Association, A.D. (2004). Physical activity/exercise and diabetes. Diabetes care, 27(suppl 1):s58-s62.

26. La Torre, G., Sferrazza, A., Gualano, M. R., De Waure, C., Clemente, G., De Rose, A. M., & Ricciardi, W. (2014). Investigating the synergistic interaction of diabetes, tobacco smoking, alcohol consumption, and hypercholesterolemia on the risk of pancreatic cancer: a case-control study in Italy. BioMed research international.

27. Allcock, D. M., & Sowers, J. R. (2010). Best strategies for hypertension management in type 2 diabetes and obesity. Current diabetes reports, 10(2), 139-144.