ISSN 1016—2232

THE Dhaka University Studies

PART E
(BIOLOGICAL SCIENCES)

VOL. 5 NO. 2

JULY 1990

Dhaka University Studies. Part E 5(2) : 137-150, 1990 (July)

DETECTION OF MULTICOLLINEARITY AND IDENTIFICATION OF RISK FACTORS OF *DIABETES MELLITUS*

NUSRAT JAHAN[1] AND M. ATAHARUL ISLAM

Department of Statistics, University of Dhaka, Dhaka 1000, Bangladesh.

Abstract

In addition to insulin deficiency as the major cause, socioeconomic factors, environmental factors, food habit and similar factors also contribute to the development of diabetes mellitus. This study is intended to identify contribution of these factors to diabetes mellitus by using multivariate analysis. The factors which are found to have positive relationship with blood glucose level are acetone in urine, albumin in urine, erythrocytic sedimentation rate, white cell count in blood, and pulse rate. The factors having negative relationship with blood glucose level are body mass index adjusted for sex, education, systolic blood pressure, place of residence, marital status, age and heredity. Findings of this study may provide useful guidelines to the policy makers, researchers, and health practitioners.

Introduction

Large number of people from all socioeconomic levels of the globe are affected by *diabetes mellitus*, It is recognised by chronic elevation of blood glucose level (1). The symptoms of the disease are severe thirst, profuse urination, weight loss, weakness etc. (1). Although it is known that the major cause is the deficiency of insulin, however, other socioeconomic factors such as education, food habit, nutritional status, biomedical conditions and other factors enhance diabetic state (1). Factors such as obesity, malnutrition, change of life style etc. are also identified as risk factors by World Health Organisation (WHO). In some studies, age and sex (2), and

[1]To whom correspondence and reprint reguests should be addressed.

Abbreviations : Fasting blood glucose level (FBG), blood glucose level 2 h after 75 g of glucose load (TBG), age, sex, marital status (MAR), education (EDU), area of residence (AREA), type of work (TYWR), annual income of the family (ANIF), family history of father (FATH), family history of husband or wife (HW), betel habit (BET), heredity, (HED), albumin in urine (ALBU), acetone in urine (ACET), body mass index adjusted for sex (PBMI), erythrocytic sedimentation rate (ESR), white cell count (WHCL), systolic blood pressure (SB), diastolic blood pressure (DB), lymphocytes (LYMP), polymorphonuclear leukocytes (POLY), pulse rate (PL), treatment after registration in BIRDEM (TRAF).

heredity (3) are also found as risk factors. In addition, the presence of albumin (4) and acetone (5) in urine give indication of association with *diabetes mellitus*.

The objective of this study is to identify the risk and prognostic factors of *diabetes mellitus* with the help of multivariate analysis. The relative importance of each independent variable on the dependent variable can be ascertained in this type of analysis. Since there are many variables that contribute in the development of the discase such multivariate analysis may provide very useful information to medical practitioners and health policy makers.

Methodology

This study is based on the data of *diabetes mellitus* patients registered at BIRDEM (Bangladesh Institute of Research and Rehabilitation in Diabetes, Endocrine and Metabolic Disorders) in 1984. The total number of observations were 1625. In BIRDEM for each patient, information regarding patients' demographic characterstics, socioeconomic characteristics, family history, habits, past medical history, diet, clinical features, systematic medical examination, glucose tolerance tests etc. are recorded. The selected variables for this study are, fasting blood glucose level (FBG), blood glucose level 2 h after 75 g of glucose load (TBG), age, sex, marital status (MAR), education (EDU), area of residence (AREA), type of work (TYWR), annaul income of the family (ANIF), family history of father (FATH), family history of husband or wife (HW), betel habit (BET), heredity (HED), albumin in urine (ALBU), acetone in urine (ACET), body mass index[2] adjusted for sex (PBMI), erythrocytic sedimentation rate (ESR), white cell count (WHCL), systolic blood pressure (SB), diastolic blood pressure (DB), lymphocytes (LYMP), polymorphonuclear leukocytes (POLY), pulse rate (PL), treatment after registration in BIRDEM (TRAF).

The standard regression model can be written as,

$$Y = \beta o + \beta_1 X_1 + \ldots \ldots \ldots + \beta_{k-1} X_{k-1} + e$$

where Y is the dependent variable and X_i's, i = 1, 2, ..., k–1 are independent variables. βo is called the regression constant and β_i's are known as regression coefficients, e's are called the error terms (6).

[2]Body mass index BMI=(weight) in kg/square of height in meter, For male patients, PBMI=(BMI/22.1 ×100 and for female patients PBMI=(BMI/20.6)×100.

In this study two multiple regression techniques will be applied to the data of *diabetes mellitus*. These are all possible equations and stepwise methods. The reason for including the all possible method is to find whether the stepwise regression method provides us with the similar findings or not. Among all possible equations the best equation is selected after comparing all the equations on the basis of R^2, adjusted R^2, residual mean square and Mallow's C_p statistic.

To perform regression analysis on the data of the patients, it is necessary to identify the dependent and independent variables. Blood glucose level is considered as the diagnostic criteria for the disease in this study (7). According to the official diagnostic criteria given by WHO in 1985, a person is diagnosed as a diabetic if either his fasting-blood glucose level is greater than 140 mg dl^{-1} or blood glucose level 2-h after 75 g of glucose load is greater than 200 mg dl^{-1}. Of the two criteria, the latter is preferred as the fasting state can rarely be assured. In this study the blood glucose level 2h after 75g of glucose ingestion, is considered as the dependent variable.

After primary selection of independent variables on the basis of correlation and χ-square test, these variables are further investigated to detect multicollinearity among them. For detecting multicollinearity, the method suggested by Belsley *et al.* (8) is used in this study.

Results and Discussion

After checking the correlation and association among independent variables the following variables are primarily selected as indepedent variables for the dependent variable, blood glucose level 2h after 75g of glucose load (TBG). These are age, AREA, MAR, EDU, TYWR, ANIF, FATH, HW, PL, HED, ALBU, ACET, PBMI, ESR, WHCL, SB, DB, LYMP, POLY, TRAF. All the categorical variables of this study except EDU are dichotomous, and therefore, can be represented by one dummy variable in the analysis. In case of education, low and medium (EDU 1), and high (EDU 2) level of education are treated as three distinct categories and the analysis is performed accordingly.

After primary selection of independent variables, these are again checked for multicollinearity. In the following table the condition indexes[3] and associated variance decomposition for each of the independent variable corresponding to the dependent variable blood glucose level–2h after 75g of glucose ingestion are given (Table 1).

[3]Condition index is defined as the square root of the ratio of the largest eigenvalue to each of remaining eigenvalues. Each eigenvalue is numbered in order of magnitude. Twenty one variance decomposition proportions are obtained for twenty independent variables and the intercept.

Table 1. Results of collinearity diagnostics for blood glucose level (two hours) with corresponding independent variables for diabetes mellitus patients of BIRDEM, 1984.

Number	Eigen Value	Condition Number	Variance Proportion									
			Intercept	AGE	ESR	WHCL	SB	DB	LYMP	POLY	FATH	MAR
1	14.4465	1.0000	0.0000	.0002	.0010	.0002	.0000	.0000	.0001	.0000	.0005	.0005
2	1.2367	3.4178	0.0000	.0001	.0056	.0001	.0000	.0000	.0000	.0000	.1883	.0000
3	1.0407	3.7257	0.0000	.0000	.0031	.0000	.0000	.0000	.0000	.0000	.0692	.0006
4	0.9739	3.8514	0.0000	.0001	.0023	.0000	.0000	.0000	.0000	.0000	.0815	.0007
5	0.8633	4.0908	0.0000	.0000	.0046	.0000	.0000	.0000	.0000	.0000	.0048	.0001
6	0.5252	5.2444	0.0000	.0001	.0150	.0000	.0000	.0000	.0000	.0000	.0568	.0001
7	0.4619	5·5925	0.0000	.0011	.0116	.0001	.0000	.0000	.0001	.0000	.0001	.0032
8	0.3629	6.3097	0.0000	.0000	.0369	.0000	.0000	.0000	.0000	.0000	.5462	.0078
9	0.3297	6.6196	0.0000	.0010	.6927	.0002	.0000	.0000	.0001	.0000	.0185	.0030
10	0.1903	8.7121	0.0001	.0064	.0261	.0045	.0004	.0003	.0001	.0002	.0020	.0213
11	0.1644	9.3736	0.0001	.0002	.0177	.0024	.0001	.0001	.0001	.0002	.0032	.0252
12	0.1077	11.5793	0.0001	.0245	.0898	.0024	.0050	.0020	.0045	.0001	.0197	.6431
13	0.0771	13.6858	0.0000	.0356	.0081	.0003	.0005	.0001	.0017	.0004	.0030	.1921
14	0.0646	14.9506	0.0000	.0089	.0439	.2726	.0004	.0001	.0696	.0114	.0003	.0433
15	0.0533	16.4582	0.0001	.3503	.0100	.2401	.0025	.0001	.0445	.0026	.0003	.0028
16	0.0379	19.5249	0.0000	.3528	.0090	.1986	.0174	.0164	.0249	.0005	.0017	.0009
17	0.0267	23.2625	0.0000	.0964	.0009	.0031	.0827	.0410	.0002	.0003	.0011	.0023
18	0.0212	26.0731	0.0067	.0461	.0074	.2162	.0619	.0154	.0125	.0496	.0021	.0292
19	0.0101	37.7482	0.0112	.0090	.0092	.0257	.0007	.0025	.0641	.1050	.0001	.0049
20	0.0045	56.5713	0.0008	.0590	.0045	.0009	.8273	.9196	.0041	.0074	.0004	.0137
21	0.0011	114.620	0.9809	.0081	.0008	.0010	.0023	.7734	.8177	.8177	.0001	.0052

Table 1. (Continued)

Number	Variance		Proportion								
	AREA	ANIF	PL	HED	TYWR	PBMI	EDU 1	EDU 2	ALBU	ACET	TRAF
1	.0007	.0007	.0001	.0009	.0011	.0001	.0004	.0003	.0005	.0001	.0011
2	.0008	.0006	.0000	.0734	.0043	.0000	.0102	.0253	.0403	.0097	.0124
3	.0018	.0001	.0000	.0366	.0219	.0000	.0193	.0498	.0003	.2422	.0013
4	.0007	.0008	.0000	.0284	.0008	.0001	.0023	.0126	.0477	.6057	.0022
5	.0002	.0000	.0000	.0000	.0125	.0000	.0051	.0063	.7453	.0675	.0056
6	.0068	.0044	.0000	.0088	.0633	.0005	.0019	.0007	.0068	.0446	.6753
7	.0015	.0061	.0001	.0003	.7738	.0002	.0271	.0090	.0054	.0044	.0314
8	.0001	.0005	.0000	.7213	.0249	.0001	.0036	.0000	.0013	.0012	.0187
9	.0044	.0095	.0000	.0438	.0225	.0000	.0049	.0009	.1032	.0004	.1333
10	.8444	.0065	.0004	.0159	.0010	.0000	.0000	.0081	.0015	.0044	.0194
11	.0507	.9080	.0003	.0353	.0027	.0001	.0023	.0010	.0008	.0046	.0067
12	.0011	.0205	.0020	.0242	.0110	.0071	.0551	.0331	.0009	.0016	.0005
13	.0194	.0106	.0004	.0019	.0055	.0000	.7577	.9129	.0042	.0009	.0001
14	.0075	.0087	.0000	.0001	.0000	.0043	.0032	.0012	.0001	.0001	.0100
15	.0080	.0015	.0007	.0028	.0225	.0071	.0685	.0766	.0010	.0072	.0153
16	.0358	.0029	.0020	.0032	.0292	.2124	.0050	.0127	.0015	.0001	.0046
17	.0101	.0086	.0031	.0000	.0004	.6425	.0002	.0020	.0146	.0001	.0191
18	.0036	.0061	.1643	.0007	.0004	.1046	.0096	.0150	.0214	.0044	.0336
19	.0007	.0014	.7919	.0004	.0005	.0167	.0152	.0233	.0002	.0000	.0021
20	.0015	.0008	.0000	.0013	.0000	.0039	.0009	.0028	.0018	.0006	.0009
21	.0000	.0016	.0346	.0009	.0019	.0002	.0076	.0063	.0014	.0003	.0064

In Table 1, 3 condition indexes are found to be greater than 30. The highest one is 114.620, which is associated with large variance decomposition proportions of polymorphonuclear leukocytes and lymphocytes. The next highest condition index 56.571 is associated with large variance decomposition proportions of systolic and diastolic blood pressure. These two instances indicate strong linear dependencies among respective variables. The third highest condition index is 37.748 but it is not associated with large variance decomposition proportion of at least two variables which is essential for diagnosing collinearity. Therefore, collinearity can be diagnosed only in the first two cases. Hence, polymorphonuclear leukocytes and diastolic blood pressure are excluded from the set of independent variables to be used in the regression analysis.

For evaluation, all possible equations obtained are divided into different sets depending on the number of variables in the equation, e.g., all equations containing one independent variables constitute a set. Similarly all equations containing two variables constitute another set and so forth. Then in each set, equations are compared on the basis of R^2, adjusted R^2, residual mean squares and Mallow's C_p statistic. The equation with the largest R^2 value, the smallest residual mean square and C_p value equal to or close to the number of independent variables in the equation is selected from each set. Then among the leading equations of different sets comparisons are made and one equation is selected as the resultant equation. Not necessarily the equation with the largests R^2 value is selected, as the equation containing all the variables have the largest R^2 value. But it may happen that after certain variables are entered in the equation, the increase in R^2 may become negligible and C_p value may closely approximate the number of independent variables in the equation.

To evaluate all possible equations for this study, the command RSREG in Statistical Analysis System (SAS) package program is used. For the evaluation of stepwise regression, REGRESSIGN subcommand with option STEPWISE in Statistical Package Program for Social Science (SPSS) is used.

The results for the dependent variable TBG, obtained by applying all possible and stepwise methods to the same set of independent variables are given in the Tables 2 and 3, and are referred to as model (a).1 and (a).2, respectively.

The 18 independent variables are, AGE, PBMI, ESR, WHCL, SB, LYMP, PL, AREA, EDU1, EDU2, MAR, TYWR, ANIF, HED, FATH, ALBU,

ACET, and TRAF. After evaluating all possible equations for model (a).1, an equation containing 15 variables have been selected as the best equation. The 15 variables are TRAF, ACET, ALBU, PBMI, SEX, EDU2, SB, HED, EDU1, AREA, AGE, ESR, MAR, BET, and LYMP. The estimated coefficients and their corresponding standard errors for this equation is given in Table 2.

Table 2. Regression estimates of coefficients for model (a).1 by using all possible methods (BIRDEM, 1884).

Variable	Estimated coefficient (b)	S.E. (b)	T	Sig T
LYMP	—.2605	.2079	—1.253	.2100
EDU1	—10.3656	5.7030	—1.818	.0693
SB	—.2308	.0793	—2.909	.0037
ACET	51.0994	9.7939	5.217	.0000
HED	—5.2436	3.4152	—1.535	.1249
MAR	—15.8152	4.9786	—3.177	.0015
AREA	—8.3382	3.9855	—2.092	.0343
TRAF	71.4084	3.2862	21.727	.0000
ALBU	20.5695	5.0025	4.112	.0000
ANIF	—4.4692	.1483	—1.077	.2815
AGE	—.5797	.1410	—4.110	.0000
ESR	.1738	.0734	2.368	.0100
PBMI	—.2884	.0796	—3.623	.0003
EDU2	—25.0987	6.2083	—4.043	.0001
Constant	360.2888	19.3178	18.651	.0000

R^2 : 0.3632 (p=0.000), Adjusted R^2 : 0.3573

For all possible regression equation model, 15 independent variables are found to have significant effect on blood glucose level 2 h after 75 g of glucose load (TBG). Of these only 5 variables have positive effect on TBG. The variables are TRAF, ACET, ESR, PL, and ALBU. The remaining 10 variables PBMI, AGE, EDU1, ANIF, HED, AGE, EDU2, AREA MAR, and SB have negative effect on TBG.

Table 3 displays results obtained by stepwise method for the dependent variable TBG.

For stepwise method, 11 independent variables are found to contribute significantly on TBG. Of these variables, TRAF, ACET, ALBU, and ESR have positive effect and EDU2, PBMI, AGE, MAR, SB, AREA, and EDUI have negative effect on the dependent variable TBG.

Table 3. Regression estimates of coefficients for model (a).2 by using stepwise methods (BIRDEM, 1984).

Variable	Estimated coefficient (b)	S.E. (b)	T	Sig T
TRAF	72.0854	3.2795	21.981	.0000
ACET	51.8316	9.7908	5.294	.0000
EDU2	—27.1281	6.1413	—4.417	.0000
PBMI	—.3228	.0776	—4.156	.0000
ALBU	20.8423	5.0012	4.167	.0000
AGE	—.5604	.1402	—3.998	.0001
MAR	—14.6455	4.9501	—2 959	.0031
SB	—.2267	.0783	—2.893	.0039
ESR	.1982	.0718	2.758	.0059
AREA	—9.1325	3.9673	—2.302	.0215
EDU1	—12.0004	5.6651	—2.118	.0343
Constant	366.8695	14.1757	25.880	.0000

R^2 : 0.3601 (p=0.000), Adjusted R^2 : 0.3358

Model (a).2 has smaller R^2 value than model (a).1. Overall fit for both the all possible and stepwise regression models are highly significant ($p<0.001$).

The factor TRAF i.e., treatment after registration is included in the models because it gives an indirect measure of patients' insulin deficiency. However, some ambiguity arises for inclusion of the factor TRAF because partly it represents the dependent variable. That is why in Tables 4 and 5 all possible and stepwise regression methods are again applied to the same set of independent variables excluding TRAF and the models are referred to as (b).1 and (b).2 respectively.

Table 4. Regression estimates of coefficients for model (b).1 by using all possible methods (BIRDEM, 1984).

Variable	Estimated coefficient (b)	S.E. (b)	T	Sig T
LYMP	—.3099	.2406	—1.288	.1979
EDU1	—13.8118	6.4809	—2.131	.0332
SB	—.2821	.0901	—3.130	.0018
ACET	65.4554	11.1104	5.891	.0000
HED	—10.1462	3.8734	—2.619	.0089
MAR	—17.9443	5.6588	—3.171	.0015
AREA	—12.1339	4.5353	—2.675	.0075
PL	0.3679	.1892	1.945	.0520
WHCL	0.0010	8.9151E-04	1.163	.2452
ALBU	19.5117	5.6944	3.426	.0006
ANIF	—5.2999	4.7166	—1.124	.2613
AGE	—.8620	.1604	—5.374	.0000
PBMI	—.6068	.0889	—6.820	.0000
ESR	.3176	.0835	3.801	.0001
EDU2	—34.5750	7.0408	—3.911	.0003
Constant	437.4864	23.7464	18.423	.0000

R^2 : 0.1771 (p=0.000) Adjusted R^2 : 0.1694

For model (b).1, 15 independent variables are found to have significant effect on blood glubose level 2 h after 75 g of glucose load (TBG). Of these only 5 variables have positive effect on TBG. The variablos are ACET, ESR, PL, ALBU, and WHCL. The remaining 10 vartables PBMI, AGE, EDU1, ANIF, HED, MAR, EDU2, AREA, LYMP, and SB have negative effect on TBG.

For stepwise method in addition to TRAF, 15 outlying observations are also deleted from the analysis.

Table 5. Regression estimates of conficients for madel (b).2 by using stepwisn methods (BIRDEM, 1985),

Variable	Estimated coefficient (b)	S.E. (b) Standard Error (b)	T	Sig T
PBMI	—.6145	.0845	—7.274	.0000
ACET	78.1926	10.9484	7.142	.0000
EDU2	—20.9877	6.1413	—4.417	.0000
ESR	.3517	.0718	2.758	.0059
AGE	—.8839	.1402	—3.998	0001
AREA	—14.1579	3.9673	—2.302	0215
ALBU	21.4122	5.0012	4.167	.0000
MAR	—20.7227	4.9501	—2.959	.0031
HED	—11.0876	3.7125	—2.987	.0029
SB	—.2780	.0783	—2.893	.0039
PL	.5209	.1836	2.837	.0046
WHCL	.0017	8.5127E-04	—2.085	.0372
Constant	394.4270	20.7766	18.984	.0000

R^2 : 0.1892 (p=0.000), Adjusted R^2 : 0.1831

For model (b).2, 12 independent variables are found to contribute significantly on TBG. Of these variables ACET, ALBU, ESR, WHCL, and PL have positive effect and EDU2, PBMI, AGE, MAR, SB, AREA, and HED have negative effect on the dependent variable TBG.

It should be noted that after TRAF has been excluded from the analysis the value of R^2 reduced almost to half in both the all possible and stepwise regression models. Although the overall fit for both the models are highly significant ($p<0.001$).

Conclusion and Implication of the Study

The main objective of this study is to find the best possible regression equation that identifies the determinants of diabetes mellitus. For this purpose the blood glucose level 2 h after 75g of glucose load is considered as the

dependent variable in this study. Independent variables are selected primarily on the basis of the relationship they have with the dependent variables as well as with themselves. Then for further investigation of whether multicollinearity exists or not the method suggested by Belsly *et al.* (8) is employed here. By using their method it has been observed that the condition indexes relating to polymorphonuclear leukocytes and lymphocytes, and systolic and diastolic blood pressure are greater than 30 and these are associated with high variance decomposition proportion (Table 1) which indicate strong collinearity. To remove the effects of collinearity, the varialbes polymorphonuclear leukocytes and diastolic blood pressure are excluded from further analysis.

In model (a).1 the number of independent variables is 18. For this model 11 variables are identified as significant by all possible and stepwise methods. Of these 11 variables, treatment after registration in BIRDEM, albumin, acetone and ESR have positive effect on 2 h blood glucose level. The remaining 7 variables, higher education, BMI adjusted for sex, age, marital status, systolic blood pressure, area of residence, and medium education have negative effect on 2 h blood glucose level. All of these 11 variables are significant at 5% level. All possible method identified additional 4 variables. These are pulse rate, lymphocytes, heredity, and annual income of the family. Of these 4 variables only pulse rate has positive effect on 2 h blood glucose level. All these variables have been found to be nonsignificant at 5% level of significance.

The factor treatment after registration in BIRDEM is included in the analysis because it gives an indirect measure of patients' insulin deficiency. However some ambiguity arises for its inclusion because to some extent it represents the dependent variable. That is why in Tables 4 and 5 the results are obtained by applying all possible and stepwise methods respectively after excluding the said factor.

After the exclusion of the factor treatment after registration in BIRDEM from the analysis, the number of independent variables become 17. The models obtained by applying the all possible and stepwise methods are referred to as models (b).1 and (b).2, respectively. In both the models (b). 1 and (b). 2, 12 variables are found to exert significant effect (Tables 4 and 5) on 2 h blood glucose level. Of them, acetone, albumin, ESR, pulse rate, and white cell count have positive effect on blood glucose level. The remaining 7 variables i.e., age, marital status, area, heredity, higher education, body

mass index adjusted for sex, and systolic blood pressure exert negative effect on the blood glucose level. The stepwise method identified only the above mentioned 12 variables. But the all possible model i.e.. model (b).1 identified 3 additional variables. These are lymphocytes, medium education, and annual income of the family. All of them effect negatively on the development of 2 h blood glucose level.

After the exclusion of the factor treatment after registration from analysis, white cell count, and pulse rate are found to exert positive effect on 2h blood glucose level by both the all possible method and stepwise method. It should be noted that the value of R^2 has been reduced almost to half in both the cases. But for models (b).1 and (b).2, the value of R^2 is greater in the stepwise method than in the all possible method (Tables 4 and 5) whereas in case of models (a).1 and (a).2. the all possible method gave greater R^2 value (Tables 2 and 3).

After comparing the results obtained by applying two different multiple regression methods for two cases, no substantial variation in R^2 is observed. In terms of both precision and economy, the stepwise regression model provides information as good as obtained by the all possible regression model.

Treatment after registration in BIRDEM gives an indirect measure of insulin deficiency due to some biochemical abnormalities for which no data are available. But this biochemical abnormality is the major cause of diabetes mellitus. The greater the biochemical abnormality the higher is the blood glucose level. This explains the large and positive effect of treatment after registration in BIRDEM.

The variable acetone in urine is also found significant with large and positive coefficient for all the models. A study conducted by Reichard *et al.* (5) confirmed the finding of high plasma acetone concentration in decompensated diabetic patients. They found that part of the acetone produced by patients in diabetic ketoacidosis was converted into glucose, thus elevating the blood glucose level further. This justifies the positive coefficient of acetone. Acetone is a prognostic factor of diabetes mellitus.

Body mass index adjusted for sex has appeared to be significant in all the models. High carbohydrate diets composed primarily of simple sugar results in exaggerated glucose intolerance. Davies *et al.* (9) suggested that maturity-onset diabetes can be controlled by monitoring fasting blood glucose level and keeping the patients to their diet. From the computational

formula of body mass index it can be seen that as weight increases, BMI should also increase. But in the all possible and stepwise methods BMI adjusted for sex has negative coefficient. This may be attributed to the continuous scale that is used to represent the BMI. In Bangladesh most of the patients are in the categories of lean and normal with only a very small proportion of obese diabetics (10). Hence the overall result in this case might be influenced by the increase in body mass index from lean to normal which is expected to have negative relationship with blood glucose level.

The variable systolic blood pressure appeared significant in all models under consideration with negative coefficient. It implies that as the systolic blood pressure increases, the blood glucose level decreases. The prevalence of hypertension in diabetic patients is known to be greater than that in non-diabetic subjects (11), however lowering blood pressure has been observed in patients with diabetic nephropathy (12). The role of systolie blood pressure on the level of blood glucose requires further examination in case of Bangladeshi patients.

Albumin in urine appeared positively significant in all the models. This means that the greater the quantity of albumin found in urine, the higher the blood glucose level will be. This is also supported in a study by Viberti *et. al.*(4). They found that albumin excretion has significantly elevated in the diabetics than non-diabetic patients.

Education appeared significant in all the models. It has a negative coefficient i.e., education has negative relationship with blood glucose level. This indicates that diabetics with higher level of education are more aware about the disease. Nevertheless, educated persons contact with doctors at relatively early stages of the development of the disease.

Place of resisdence is also found to have significant association with blood glucose level. The association is negative in all the models. Place of residence has been included in the analysis by coding 1 for all patients living in urban areas and 0 for living in rural areas. Hence the negative coefficient suggests that people living in urban areas have relatively lower blood glucose level than the people in rural areas. This is indicative of the fact that the people in urban areas avail the facilities of detection and treatment of diabetes mellitus at an earlier stage than those living in rural areas.

Heredity also appeared significant in some of the models. The estimated coefficient of heredity is negative, implying that heredity has inverse effect on

the 2 h blood glucose level. The etiology of diabetes mellitus in terms of inheritance of the disease is still unclear in context to Bangladesh. The negative relationship may be due to the fact that the individuals from the families having any other diabetic member in the family are more conscious about the incidence of the disease and they are more careful about the control of the disease.

Marital status has negative relattionship with blood glucose level. Marital status is incorporated in the analysis by coding 1 for all currently married patients and 0 for all currently not married patients. Hence, the negative coefficient implies that currently married persons are less likely to have elevated blood glucose level than persons currently not married.

In all the models, ESR is found to have positive relationship with the blood glucose level, indicating that with increase in the erythrocytic sedimentation rate the blood glucose level also increases.

Age has negative relationship with the bloood glucose level, that is, young people are more likely to have elevated blood glucose level than older people. But majority of the diabetic patients in our country have non insulin dependent diabetes mellitus (NIDDM), which usually occur among adults (13). This study includes only those persons who registered at BIRDEM for the first time. A possible explanation of elevated blood glucose level among young people is that an individual at younger age registers at BIRDEM only if the person has relatively severe diabetes.

White cell count has positive relationship with the blood glucose level. It has appeared significant in the models (b).1 and (b).2 i.e. only after the exclusion of treatment after registration from analysis.

After analyzing the results obtained by different multiple regression methods, it can be concluded that the stepwise method can be employed conveniently as compared to all possible method in terms of both economy and precision. This paper makes an attempt to solve the problem of multicollinearity in the field of *diabetes mellitus*. In selecting a good model for analayzing data, a comparison is made between the all possible and the stepwise methods. Considering economy, time, and precision, the stepwise method has been found to be a convenient method from this analysis. Serveral variables have been identified as exerting significant effect on the blood glucose level which may provide useful guideline to the policy makers, researchers and health practitioners.

Acknowledgements

We thank the Diabetic Association of Bangladesh for allowing us to use their data and library facilities for the purpose of this study. In this regard we thank the late national Professor Mohammad Ibrahim and Professor Azad Khan of Diabetic Association. For the part of analysis of this paper we have used extensively the Computer Center of the University of Dhaka. We are particularly grateful to Mahmood Hasan, systems analyst of Computer Center. We also thank International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR'B) for allowing us to use their computer facilities.

References

1. World Health Organisation 1985. Report of WHO Study Group; Technical Report Series 727. pp. 9-11. WHO, Genea.
2. Ibrahim, M. 1962. Diabetes in East Pakistan. British Medical J. **2** : 837-839.
3. Gleslie, R. D. 1986. Metabolic abnormalities in children of noninsulin dependent diabetics. British Medical J. **23** : 840-842.
4. Viberti, G.C., Pickup, J.C., Jarrett, R.J., and Keen, H. 1979. Effects of control of blood glucose on urinary excretion of albumin and microbuli in IDDM. New England J. of Medicine, **300** : 638-641.
5. Reichard, G.A. Jr., Skutches, C.L., Hoeldtke, R.D., and Owen, O.E. 1986. Acetone metabolism in humans during diabetic ketocidosis. Diabetes, **35** : 668-674.
6. Draper, N.R. and Smith, H. 1966. Applied Regression Analysis. pp. 17, 70-85. Wiley, New York.
7. Keen, H. 1986. The diagnosing and classification of diabetes mellitus. In : *World Book of Diabetes in Practice 1986* (Krall, L.P., Alberti, K.G.M.M., and Turtle, J.R. eds.). p. 9. Exercepta Medica, Elsevier, Amsterdam.
8. Belsley, D.A., Kuh, E., and Welsch, R.E. 1980. Regression Diagnostics : Identifying Influential Data and sources of Collinearity. pp. 96-115. Wily, New York.
9. Davies, S.H., Simpson, R.W., and Turner, R.C. 1980. Control of maturity-onset diabetes by monitoring fasting blood glucose and body weight. Diabetes Care, **3** : 607-610.
10. Ibrahim, M. 1960. Tropical Diabetes. Published in the IVth Pan Pacific Ocean Science Congress, Sec. G, Human Ecology. p. 189.
11. Pell, S. and D' Alonzo, C.A. 1967. Some aspects of hypertension in diabetes mellitus. J. of American Medical Assoc. **202** : 104-110.
12. Bjork, S., Nyberg, G., Mulec, H., Granerus, G., Herlitz, H., and Aurell, M. 1986. Beneficial effects of angiotensin converting enzyme inhibition on renal patients with diabetic nephropathy. British Medical J. **293** : 471.
13. Karem, J.H. 1985. Diabetes mellitus, hypoglycemia and lipoprotein disorders. In : *Current Medical Diagnosis and Treatment* (Krupp, M.A., Chatton, M.J., and Werdeger, D. eds.). pp. 761-792. Lange Medical Publications, New York.

(Manuscript received on 18.1.1990; revised on 30.6.1990)