EasyChair Preprint

№ 9513

Diabetes Prevalence in Developing Countries

Tahir Farooq and Zohaib Sohail

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

January 3, 2023

Diabetes Prevalence in Developing Countries

Tahir Farooq
Department of Economics
Information Technology University
Lahore, Pakistan
bseds19004@itu.edu.pk

Zohaib Sohail
Department of Economics
Information Technology University
Lahore, Pakistan
bseds19011@itu.edu.pk

***Abstract*—Prevalence of diabetes has been highlighted as being very fast paced, especially in the developing countries. In many countries it has been considered as the main cause of death. According to the 10th edition of International Diabetes Federation (IDF), 537 million people are living with diabetes, and it has been responsible for 6.7 million deaths in 2021 alone, which is 1 person every 5 seconds. The main health related causes underlined for the rise in diabetes prevalence include factors such as overweight, obesity, insufficient physical activity, hypertension, and unavailability of efficient health services in a country. Many other factors falling under the umbrella of socio-economic factors also play a role of equal importance in diabetes prevalence. However, for the purpose of our paper, we would be focusing upon health and lifestyle related variables along with a few expenditure variables to know the impact upon diabetes for expenditure on health care in a country. Statistical analysis upon the data of 7 countries from the year 2006-2016 would be performed to analyze the impact of independent variables upon our model dependent variable of 'Prevalence of diabetes'. The independent variables used are continuous in nature. The paper doesn't rely upon simple OLS regression, rather Fixed-Effects (FE) regression model was used to control time and country fixed effects. As the research undertaken was secondary in nature, the paper had to rely upon complete data available which was not the case for every variable that should be included for better results, hence our analysis does have some limitations.**

Keywords—diabetes, developing countries, overweight, obesity, health services, socio-economic, fixed effect model

I. Introduction (*Heading 1*)

A rapid increase in diabetes is being experienced in developing countries all over the world. According to the World Health Organization (WHO), "Diabetes is a major health burden in developing countries"[1]. Diabetes is a chronic, non-communicable disease with two major types. Type 1 diabetes is prevalent among the young population due to the absence of insulin production. Type 2 diabetes on the other hand is diagnosed among the adults due to the deficiency of insulin. Both, Type 1 and 2 diabetes can cause serious health implications, if not provided the right medical and physical care.

In the light of a report issued by the International Diabetes Federation(IDF), 246 million people are affected by diabetes worldwide. This number is expected to increase to 380 million by 2025. According to another report, "prevalence of diabetes in developing countries has increased dramatically over the past 20 years"[2]. The report also states that the majority of the population affected by the disease would be among the age of 45-65 years, primarily living in low- and middle-income countries. This research is aimed at understanding the relationship between a non-communicable disease and the economic output of a country. Additionally, it would also explore factors that are associated with high prevalence of diabetes in a developing country. Through using various panel data variables, the purpose of this research is to find out trends and summarize data associated with prevalence of diabetes in developing economies. In particular it explores the relationship of lifestyle as well as expenditure of health on the prevalence of diabetes.

The key feature of this research is to study a group of factors that influence diabetes prevalence in countries with similar socioeconomic conditions whereas the literature that previously exists only focuses on a singular factor or comparative analysis between two countries.

II. Literature review (*Heading 2*)

WHO advises its country members to reduce physical inactivity by 10% at the national level by 2025 in order to battle the rising burden of NCDs through its global monitoring framework for Noncommunicable disease (NCDs). According to a recent 16-year trend analysis, the current 3.5% decline in physical inactivity is too slow to meet this goal. Additionally, the WHO global action plan on physical activity established a target of a 15% relative decrease in adult physical inactivity from the baseline year of 2016 by 2030. To reach this goal, many countries are stepping up their efforts by strengthening surveillance, researching correlates, adopting policies, and carrying out action plans [3].

If we discuss the prevalence in developing countries, India has an estimated 8.7% prevalence in adults[4],China has an estimated 11.6% [5], while Egypt has a high prevalence of 12.3% amongst their adults[6].According to IDF, "in 2017, an estimated 425 million adults aged 18 years and older had diabetes worldwide, and this number is projected to increase to 629 million by 2045"[7].

Studies are being conducted more frequently to determine the levels of physical activity and its contributing factors globally. The process of achieving the aim is hampered by poor implementation techniques, misinformation, a lack of educated labor, and insufficient cross-sectoral cooperation. There has been a rapid epidemiological transformation in India, as seen by the estimated doubling of the disability-adjusted life years related to NCDs (from 29% in 1990 to 58% in 2019). One of the main NCD risk factors is physical inactivity. In light of this, research evaluating the incidence of adults' physical inactivity has been conducted over a number of years in a few Indian states. But there hasn't been enough research done on the factors that contribute to Indian people getting insufficient PA [3].

For the purpose of our research, we did find data for the prevalence of inactivity within a country but that was incomplete, therefore, we included a proxy variable of Hypertension in our model, as we assume in our paper that an increase in physical inactivity leads to hypertension, especially among adults. Naha S in her book ‘Hypertension in diabetes’ also sheds light upon “The coexistence of hypertension and diabetes in a large population of patients” [8], as the literature suggests hypertension itself to be another important variable for our study hence, it is included in our regression model. One of the articles suggests that, "In developing countries, the dramatic rise in diabetes prevalence has been attributed to an increase in sedentary lifestyle, poor dietary habits, and rapid urbanization"[9], hence, lifestyle indicators are of much importance in our study.

In 2008, the worldwide obesity scenario was approximately 1.4 billion adults, with death rates that were 65 percent higher than those who were underweight. “United Nations General held a meeting in September 2011 to develop a Global Action Plan for the prevention and control of non-communicable diseases by 2013-2020 and to build a WHO framework to reduce obesity by 25% by 2025 and achieve obesity rates as those of 2010” [10]. Research conducted in Cebu, Philippines, in 2005 and reported in Nutritional Diabetes in July 2013 by a team of foreign and local authors revealed a rise in the prevalence of obese and overweight young adults between the ages of 18 and 24. With the help of the study's data, the obesity rate in the Philippines was projected to be 0.5% in 2005, and it was associated with neighborhood urbanization and individual socioeconomic status levels [10]. Moreover, “Diabetes prevalence is increasing rapidly in low-and-middle-income countries, where it is largely undiagnosed”[11].

As obesity has a significant impact upon the non-communicable diseases and especially diabetes, therefore, for included Mean BMI (MeanBMI) as an independent variable in our research study. A person with a BMI greater than 30 is considered to be obese, hence an increase in MeanBMI of a country would have a positive correlation with diabetes prevalence.

III. Research Methods (*Heading 3*)

The aim of our study is to observe and analyze the possible correlation between the prevalence of Diabetes in a country with its mean BMI index, hypertension prevalence, current health expenditure and out of pocket expenditure. As our research is secondary in nature, therefore the relevant data has been collected from World Health Observatory datasets. Further, the paper focuses only upon a limited number of countries, Afghanistan, Bangladesh, India, Pakistan, Philippines, Nepal, and Sri Lanka. Our research will undertake a time series analysis with the focus on the period from 2006-2016. Initially a list of 16 variables and 8 countries were identified that could play a vital role in aiding our research, however, as our research is secondary and there was a limitation towards the data we could obtain, hence, we had to drop variables with missing values and incomplete data to make our research more precise and credible. All the variables finalized for our study are continuous in nature and there is no dummy variable included. For the reliability of our data, the source itself is enough to cite as the data has been constructed by the World Health Organization (WHO). Further, the sampling methods undertaken by WHO would also be brought into discussion.

A. Hypothesis

Our hypothesis for this paper underlines that our lifestyle variables, which include an increase in the Mean BMI index of a country along with the prevalence of hypertension, would be positively correlated with the prevalence of diabetes. Moreover, the counter variable for hypertension ‘ConHypertension’ would be negatively related with diabetes. The expenditure variables in our study, such as current health expenditure and out of pocket expenditure, would be negatively correlated with the prevalence of diabetes. Hence, our null and alternative hypothesis would be as follows:

H(0) = Lifestyle variables(MeanBMI & Hypertension) have a positive correlation with the prevalence of diabetes while expenditure variables(CHEpercapita & OutofPocketexp) have a negative correlation with the prevalence of diabetes.

H(A) = Lifestyle variables (MeanBMI & Hypertension) do not have a positive correlation with the prevalence of diabetes and expenditure variables(CHEpercapita & OutofPocketexp) do not have a negative correlation with the prevalence of diabetes.

CHEpercapita stands for ‘Current Health Expenditure per Capita’ and OutofPocketexp stands for ‘Out of Pocket Expenditure’ which showcases the expenses of civilians upon medical consumption from their own pocket.

B. Model & Modification

Initially the research started with a list of 16 variables and 8 countries but there were many modifications made to make our model more credible, precise and accurate for analysis and forming claims. Firstly, the list of variables was filtered by removing the variables which had incomplete data or missing values. Also, many variables with complete data had high correlation with other variables so we had to further modify our mode and only use the set of variables which improve our model and regression analysis.

C. Correlation Matrix

(obs=80)

	OWeight	ObesAd~s	MeanBMI	Hypert~n	ConHyp~n	StunPr~e	CHEper~a	OutOfP~p
OWeight	1.0000							
ObesAdults	0.6440	1.0000						
MeanBMI	0.5828	0.9152	1.0000					
Hypertension	0.1494	0.7017	0.8002	1.0000				
ConHyperte~n	0.1136	0.1218	0.1504	-0.2815	1.0000			
StunPrevel~e	-0.5769	-0.1037	-0.1970	0.2449	-0.3184	1.0000		
CHEpercapita	0.7837	0.3266	0.4016	-0.0298	0.2140	-0.8040	1.0000	
OutOfPocke~p	-0.5525	-0.1264	-0.0937	0.0895	0.2709	0.6670	-0.6280	1.0000

As shown in the table above, many of the variables shortlisted had a high correlation with others so we had to skip them in our final model to avoid the issue of multicollinearity. Such as the variable ‘Overweight’ (Oweight) had a 0.78 correlation with CHEpercapita, ‘Obesity among Adults’ (ObesAdults) had a very high correlation of 0.91 with MeanBMI and ‘Stunting Prevalence’ (StunPrevelance) had a high negative correlation with CHEpercapita, hence, these variables were dropped and not included in the final model.

Furthermore, one of the countries included in our dataset, Indonesia, had a skewed average value for each independent variable, when compared to the averages of other countries. Therefore, we dropped Indonesia from our model to make it more efficient.

D. Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
MeanBMI	77	22.32857	.8356934	20.7	23.9
Hypertension	77	34.66234	4.227542	27.8	42.6
ConHyperte~n	77	10.2	3.053686	3.1	15.4
CHEpercapita	77	55.28182	36.39963	12.34	151.5
OutOfPocke~p	77	63.02026	9.300369	40.95	79

Each variable has continuous values of 77 observations. The first independent lifestyle variable which also forms our base case for regression and analysis is the mean BMI index in a country (named as ‘MeanBMI’ in our analysis), which is an age standardized estimate with a mean value of 22.3 and the lowest standard deviation in the list of our variables with a value of 0.83 which depicts the spread of values around the mean.

The second variable is another age standardized measure of hypertension prevalence among adults with age above 30, it is named ‘Hypertension’ in our analysis. It has a mean value of 34.6 and a standard deviation of 4.2 which shows an increased spread of values around the mean value. The third variable is a counter variable for hypertension as it shows the prevalence of controlled hypertension among adults with its name ‘ConHypertension’ in our analysis, it has a mean value of 10.2 and a standard deviation of 3.05.

Moreover, the expenditure variables included in our research are current health expenditure per capita in dollars, named as ‘CHEpercapita’, and the out of pocket expenditure as a percentage of current health expenditure in dollars, named as ‘OutofPocketexp’ in our analysis. Both of these variables have a large standard deviation of 55.3 and 63.0, respectively. These large deviations of dataset values from the mean value of these variables can be due to the variations in spending in each country at the government and individual level alike.

E. Sampling Methods

As we undertook secondary research, therefore, there are limitations towards knowing the sampling method used for the collection of data in each variable. However, as the data is published by WHO, therefore the credibility of data cannot be questioned upon.

- MeanBMI: defined as mean body mass index (BMI) in kg/m2 of defined population. It is measured based on height and weight. The data source for measuring the mean BMI is population-based surveys.
- CHEpercapita: described as current health care spending per capita represented in the relevant currency, which is the US dollar. CHE determines the average cost of healthcare for everyone. It allows worldwide comparison by helping to understand health spending in relation to population size. It is measured by dividing the current health expenditure with the whole population.
- OutofPocketexp: defined as the percentage of out-of-pocket expenses in total current medical costs. This indicator provides an estimate of the direct out-of-pocket health spending by households in each nation. It makes an estimate of the percentage of out-of-pocket expenses that make up all current medical costs.

F. Regression Model

After the final set of modifications, the regression model consisted of seven countries and 5 variables, out of which 3 can be categorized as lifestyle variables and 2 can be categorized as expenditure variables.

A base regression was first run for the correlation between MeanBMI and diabetes prevalence without controlling for both country and time fixed effects and used simple OLS estimations. A Fixed-Effects (FE) model was then added which incorporated control for both country and time fixed effects as simple OLS estimations do not capture the differences among countries. The FE-model determines individual effects of unobserved, independent variables as constant (“fix”) over time. Within FE-models, the relationship between unobserved, independent variables and the independent variables (i.e., endogeneity) can be existent. In the first simple regression, the results were as follows:

DiabetesPr~e	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
MeanBMI	.0048389	.0025964	1.86	0.066	-.0003333	.0100111
_cons	-.0681085	.0580132	-1.17	0.244	-.1836766	.0474597

After adding control for country and time, the regression results were as follows:

DiabetesPr~e	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
MeanBMI	.0125124	.001474	8.49	0.000	.0095719	.015453
cons	-.2394475	.0329153	-7.27	0.000	-.3051118	-.1737833

It can be clearly seen that MeanBMI has a greater coefficient and is more significantly impacting diabetes prevalence after the control has been added to account for time fixed effects and to capture the differences among countries. After controlling for time and country effects and adding in the controls, MeanBMI of a country does have a statistically significant effect in that year.

After our initial base case regression to know the impact of using fixed effects, we started incorporating our other independent variables one by one. The final regression equation formed according to our model is:

Prevalence in Diabetes = $\beta 0 + \beta 1$(MeanBMI)+

$\beta 2$(Hypertension) + $\beta 3$(ConHypertension)+

$\beta 4$(CHEpercapita) + $\beta 5$(OutofPocketexp) + μt

G. Fixed Effects (fe) Regression

```
Fixed-effects (within) regression               Number of obs     =         77
Group variable: ID                              Number of groups  =          7

R-sq:                                           Obs per group:
     within  = 0.7151                                         min =         11
     between = 0.1451                                         avg =       11.0
     overall = 0.1727                                         max =         11

                                                F(5,65)           =      32.63
corr(u_i, Xb)  = -0.4429                        Prob > F          =     0.0000

------------------------------------------------------------------------------
DiabetesPreva~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
        MeanBMI |   .0127454   .0027386     4.65   0.000     .0072761    .0182148
   Hypertension |   .0001626   .0012149     0.13   0.894    -.0022637     .002589
ConHypertension |  -.0032884   .0007407    -4.44   0.000    -.0047678   -.0018091
    CHEpercapita |   .0003073   .0000584     5.27   0.000     .0001908    .0004239
  OutOfPocketExp |  -.0000595   .0001118    -0.53   0.597    -.0002828    .0001639
           _cons |  -.2299876    .052036    -4.42   0.000    -.3339106   -.1260646
----------------+-------------------------------------------------------------
        sigma_u |  .02064897
        sigma_e |  .00305645
            rho |  .97856005   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(6, 65) = 217.07                     Prob > F = 0.0000
```

Our fixed effects regression model has 77 observations and the value of R-square within 0.7151 which means that all of the included variables are able to explain 71.5% of our dependent variable which is Diabetes Prevalence. The table further shows the coefficients of each dependent variable which depict the change in Diabetes Prevalence for a unit change in the listen independent variables. Moreover, the t values and p values are used to know the significance of a particular variable in affecting the dependent variable.

As discussed in our base model case, MeanBMI has the largest effect on Diabetes Prevalence as a one unit increase in MeanBMI cause the Diabetes Prevalence to increase by 0.0174 and with a high t value of 4.65, it is a very significant variable at 95% confidence interval.

Furthermore, controlled hypertension should have a negative relation with diabetes as when an individual has his hypertension under control, their chances of being diabetic would be lesser. Our coefficient of ConHypertension depicts the same claim. With a value of -0.00328, it has a negative correlation with the prevalence of diabetes. It is also another significant variable in our study with a t value of -4.44.

The only variable which doesn't have an effect as suggested in the literature is CHEpercapita. Although it should have a negative relation with Diabetes Prevalence as when the health expenditure per capita increases, diabetic patients would also decrease. However, our results show otherwise, with a coefficient of positive 0.0003, CHEpercapita depicts a positive relation with diabetes prevalence in our data. Nevertheless, we assume that our dataset includes countries which have varying degrees of funds allocated for their health system and some countries included do have serious health infrastructure problems so an increase in health expenditure per capita may not be able to outweigh the increase of diabetes sourced by other factors. Moreover, CHE is a measure done per capita, hence, some countries in our dataset have a large population size which may have an impact upon CHEpercapita.

The coefficient of Hypertension is positive as expected and OutofPocketExp coefficient depicts a negative correlation which has also been claimed by the literature. However, these variables have a very low t-value which makes them insignificant according to our dataset and model. This means that these independent variables don't have a very impactful effect upon the dependent variable as compared to the effect which other variables have.

IV. Discussion & Analysis (*Heading 4*)

While the results show a positive correlation between lifestyle factors such as Mean BMI, hypertension and a negative relation with controlled hypertension as suggested by the literature, a varying degree of results are noted for expenditure factors. There is no fixed pattern to correlation between expenditure factors and prevalence diabetes. This is majorly because of distinct economic features each of these developing countries have. Developing countries share a number of common features that include poverty, low HDI and lack of institutions to name a few but each has a distinct magnitude and impact for every country. This can be one possible explanation for a pattern as such in the case of expenditure factors.

Expenditure factors can be better explained and understood by adding other control variables such as general availability of diabetes testing and existence of operation policy for diabetes; however, no such data was available for these that could match our timeline. There were in fact a few other variables that might have allowed the model to be better equipped and obtain more accurate results such as prevalence of physical inactivity among adults as well as prevalence of obesity among adults. These could have helped to better explain lifestyle factors however, data for these was incomplete and inconsistent.

It is important to understand that our model did not account for diverse socioeconomic conditions of all the above-mentioned countries. Hence, this acts as a major limitation for our model. In addition to this, we were unable to include factors that directly and indirectly affect HDI. Our model does not accommodate health seeking attitudes of these nations and education rates, which are believed to significantly influence prevalence of diabetes in an economy. This research is only externally valid if countries from the socio-economic group are studied otherwise the research is internally valid.

Lastly, the dataset constructed for this research uses the timeline 2006-2016 because of missing data in the succeeding years. Although this does not negate the positive correlation between the aforementioned variables and diabetes prevalence alongside the surge in diabetes over the past few years, the research is not tangential with the technological advancements happening every day in the field of medicine. If more recent data was available, the influx of technology could also be included in the research to better understand the prevalence of diabetes.

To conclude our research, we would like to acknowledge the possible discrepancies and inefficiencies in our modeling and used data. Along with that we would like to highlight that our base case regression model depicted the results which were outlined by literature. Moreover, as we kept adding more variables in our model, the analysis widened and so did the scope of explanation through the used variables, as shown by the R-square value of our model. Hence, through our research we can claim that an increase in the mean BMI index of a country and an increase in the prevalence of hypertension would have a positive impact upon the prevalence of diabetes. Also, an increase in out-of-pocket expenditure would depict a decrease in the prevalence of diabetes in a country.

References

[1] World Health Organization. (2016). Diabetes. Retrieved from http://www.who.int/

[2] Shah, A., Harari, A., Ali, M. K., & Mir, S. A. (2016). Diabetes in developing countries: Epidemiology, risk factors, prevention, and control. International Journal of Endocrinology, 2016, 1-14.

[3] Ramamoorthy, T. (2022, February 11). Findings from the national noncommunicable disease monitoring survey. Human Kinetics. https://journals.humankinetics.com/view/journals/jpah/19/3/article-p150.xml?content=fulltext

[4] Kaul, U., Prabhakaran, D., Gomber, S., & Narain, P. (2018). Prevalence of diabetes and prediabetes in India: Systematic review and meta-analysis. BMJ Open, 8(3), e019654**.**

[5] Huang, Y., Bao, W., Yang, W., Zhang, B., Wang, Y., & Zhang, C. (2018). Prevalence and trends of diabetes among adults in China, 1993-2015. Diabetes Research and Clinical Practice, 140, 57-62.

[6] Mabrouk, G., Salah, S., El-Saeid, A., El-Gendy, S., & El-Gendy, A. (2017). Prevalence and associated factors of diabetes mellitus among the adult population in Egypt. BMC Public Health, 17(1), 844.

[7] International Diabetes Federation, 2021. Retrieved from IDF Diabetes Atlas 2021 | IDF Diabetes Atlas

[8] Naha, S. (2021, August 7). *Hypertension in diabetes*. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/books/NBK279027/

[9] Higgins, M., et al. (2017). Diabetes in developing countries: a need for greater awareness and action. BMC Public Health, 17(1), 715. doi:10.1186/s12889-017-4577-3

[10] Ramos, B. (2020, June). Obesity among College Students in the Northern Philippines: Input for a National Policy and University Anti-obesity Program. ResearchGate. https://www.researchgate.net/profile/Ferdinand-Bulusan/publication/342698518_Obesity_among_College_Students_in_the_Northern_Philippines_Input_for_a_National_Policy_and_University_Anti-obesity_Program/links/5f018d14a6fdcc4ca44e6f34/Obesity-among-College-Students-in-the-Northern-Philippines-Input-for-a-National-Policy-and-University-Anti-obesity-Program.pd

[11] Lozano, R., et al. (2016). Diabetes prevalence and incidence in low- and middle-income countries: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet Diabetes & Endocrinology