Risk of diabetes depending on income and living conditions

Hanna Asipovich

2022-12-20

Introduction

According to the **World Health Organisation, about 422 million people around the world have diabetes. It is a globally spread condition and the numbers have quadrupled since 1980s. In 2016, WHO published the Global report on diabetes, which calls for action to reduce exposure to the known risk factors for type 2 diabetes and to improve access to and quality of care for people with all forms of diabetes. This analysis will look at the association of an increased diabetes risk and GNI(gross national income, Atlas method). Additionally, it will consider other contributing factors, such as access to drinking water and sanitation.

Data and its key features

Source

We use the data obtained from World Bank **DataBank webpage for the years 2020-2021. On this **Github repository, one can find extracted datasets with indicators for % of country population with diabetes, GNI by Atlas method, % of country population with access to basic sanitation, % of country population with access to drinking water.

Data Cleaning

Upon examination of the datasets, we filter the data to remove NAs. We join the dataset by the country code column, remove unnecessary columns - data and country code, year. As the most recent data originates from 2020 and 2021, we decide to use these two years without considering those as time series. Data for dependent variable (% of population with diabetes) comes from the later year of 2021.

X, Y and Z Variables

Dependent variable: % of population with diabetes. Dummy variables were created for the probability models: higher than average risk for diabetes in population (above 8% population with diabetes) is equal to 1; otherwise is 0. Indepedent variable: GNI, Atlas method. Also we carried out log transformation to create lngni. Control variables: - % of population with access to sanitation;- % of population with access to drinking water.

Summary statistics

On average almost 8% of global country population has diabetes (only in 5% of countries this figure goes below 2.1% of population). Average GNI(gross national income) is USD 14,889. 87.5% of population has access to

drinking water(with 95% having their whole need covered), and 75.08% has access to basic sanitation(with 95% having their whole need covered).

Table 1: Descriptive statistics

	Mean	Median	SD	Min	Max	P05	P95
% Population with diabetes	7.89	6.70	4.50	1.10	30.80	2.10	16.78
GNI, Atlas method	14 889.74	4740.00	21 714.64	240.00	116 540.00	608.00	60 938.00
% Population with access to drinking water	87.50	95.30	16.03	37.20	100.00	54.87	100.00
% Population with access to basic sanitation	75.08	88.21	28.87	8.91	100.00	17.68	99.97

Correlation Matrix

To assess association between dependent, independent and z-variables I produced a correlation matrix (see Appendix Figure 3). It shows that in countries with a higher risk of diabetes, there is correlation with our control variables on access to drinking water and access to basic sanitation. Based on visual information by Fig.2, which shows correlation between % of population with diabetes and GNI, we can say that the regression line is in a downward direction. Hence, in countries with higher GNI, proportion of population with diabetes is lower.

Hypothesis

Our hypothesis is that higher % of population with diabetes is associated with the income inequalities and worse life conditions (access to sanitation, drinking water).

Models

Simple linear regressions

In a simple regression analysis, we used diabetes % in population as a dependent variable. Further, we applied log transformation on GNI and used it as the independent variable. Model 1 is simple linear regression and for Model 2 we use linear regression with robust standard errors. In Model 1 R2 is only 0.008 and is low to take into account. According to Model 2 variable on GNI is statistically significant, and it gives some confidence in finding association between the variables. However, R2 stays low and, hence, other variables may influence diabetes occurence, so below we will explore other probability models.

$$Diabetes = \alpha + \beta_1(lngni)$$

	Model 1	Model 2
(Intercept)	5.4564*	3.4399
	(2.1942)	(1.8132)
lngni	0.2816	0.4367*
	(0.2507)	(0.2072)
Num.Obs.	153	153
R2	0.008	
R2 Adj.	0.002	
RMSE	4.47	4.52

* $p < 0.05$, ** $p < 0.01$

Probability models

We will look further into probability of diabetes occurence as regressed on income, access to drinking water and access to basic sanitation. Earlier we created a binary variable on higher than average risk of diabetes (if % of population with diabetes is => 8% it equals to 1, otherwise to 0). We also add to our equation two more z-variables on % of population which has access to drinking water and % of population with access to basic sanitation. As the result, we can see in the Table below that our R2 increased significantly to 0.095. We can say that our z-variables contributed to the quality of the analysis.

We also see that higher GNI has negative correlation with the risk of diabetes occurrence, at the same time the table shows that access to basic sanitation and to drinking water have positive correlation, which can be explained by high level of access to those amenities globally (75% and 100% for 95% of population.)

$$Diabetes^P = \alpha + \beta_1(lngni) + \beta_2(water) + \beta_3(sanitation)$$

Table 2: The Probability of diabetes occurence- LPM, Logit (Marg), and Probit (Marg) models

	(1) LPM	(2) Logit	(3) Logit Marg	(4) Probit	(5) Probit Marg
Constant	0.268	−1.681	−1.681	−0.838	−0.838
	(0.343)	(1.755)	(1.755)	(1.015)	(1.015)
lngni	−0.093*	−0.441*	−0.094	−0.249*	−0.088
	−0.093*	−0.441*	−0.094	−0.249*	−0.249*
	−0.093*	−0.441*	−0.441*	−0.249*	−0.088
	−0.093*	−0.441*	−0.441*	−0.249*	−0.249*
	(0.043)	(0.197)	(0.051)	(0.119)	(0.045)
	(0.043)	(0.197)	(0.051)	(0.119)	(0.119)
	(0.043)	(0.197)	(0.197)	(0.119)	(0.045)
	(0.043)	(0.197)	(0.197)	(0.119)	(0.119)
sanitation	0.005	0.020	0.004	0.013	0.004
	0.005	0.020	0.004	0.013	0.013
	0.005	0.020	0.020	0.013	0.004
	0.005	0.020	0.020	0.013	0.013
	(0.003)	(0.016)	(0.003)	(0.009)	(0.003)
	(0.003)	(0.016)	(0.003)	(0.009)	(0.009)
	(0.003)	(0.016)	(0.016)	(0.009)	(0.003)
	(0.003)	(0.016)	(0.016)	(0.009)	(0.009)
water	0.007	0.039	0.008	0.020	0.007
	0.007	0.039	0.008	0.020	0.020
	0.007	0.039	0.039	0.020	0.007
	0.007	0.039	0.039	0.020	0.020
	(0.006)	(0.031)	(0.007)	(0.018)	(0.007)
	(0.006)	(0.031)	(0.007)	(0.018)	(0.018)
	(0.006)	(0.031)	(0.031)	(0.018)	(0.007)
	(0.006)	(0.031)	(0.031)	(0.018)	(0.018)
Num.Obs.	153	153	153	153	153
R2	0.095				
RMSE	0.46	0.46	0.46	0.46	0.46

* $p < 0.05$, ** $p < 0.01$

Robustness Check

To check robustness we ran logit and probit regressions on our LPM model. These models increased the coefficients and they are not exactly the same as our LPM model.

Our results, apart from association with GNI, are statistically insignificant and are not enough to say whether it is likely for a resident of a certain country to develop diabetes. Adding z-variables helped with increasing our R2. But we understand that it is worth exploring additional variables. Especially, since we know, for example, that 95% of population has their need in drinking water fully covered, we can say that this is not a telling variable for us.

For taking account external validity, we can run similar exercise on the earlier time series by the World Bank, as the research on diabetes is carried out every decade. Apart from that, we can confirm by independent national reports with more extended data. For example, see for this **report from Korea, that lower income is associated with a higher prevalence of diabetes.

Summary and conclusion

We studied the relationship between diabetes, GNI and access to drinking water and basic sanitation in countries. We built a multiple linear regression model, where we reach a conclusion that "A higher risk of diabetes occurrence has negative association with GNI level. According to the WHO 2016 report on diabetes, the prevalence of diabetes is growing especially rapidly in low and middle income countries. So it may be necessary to study more indepth what other factors are associated with the diabetes development to make an actionable proposition to the governments in this specific group of countries.

Appendix

Figure 1: Distribution of countries by % of population with diabetes

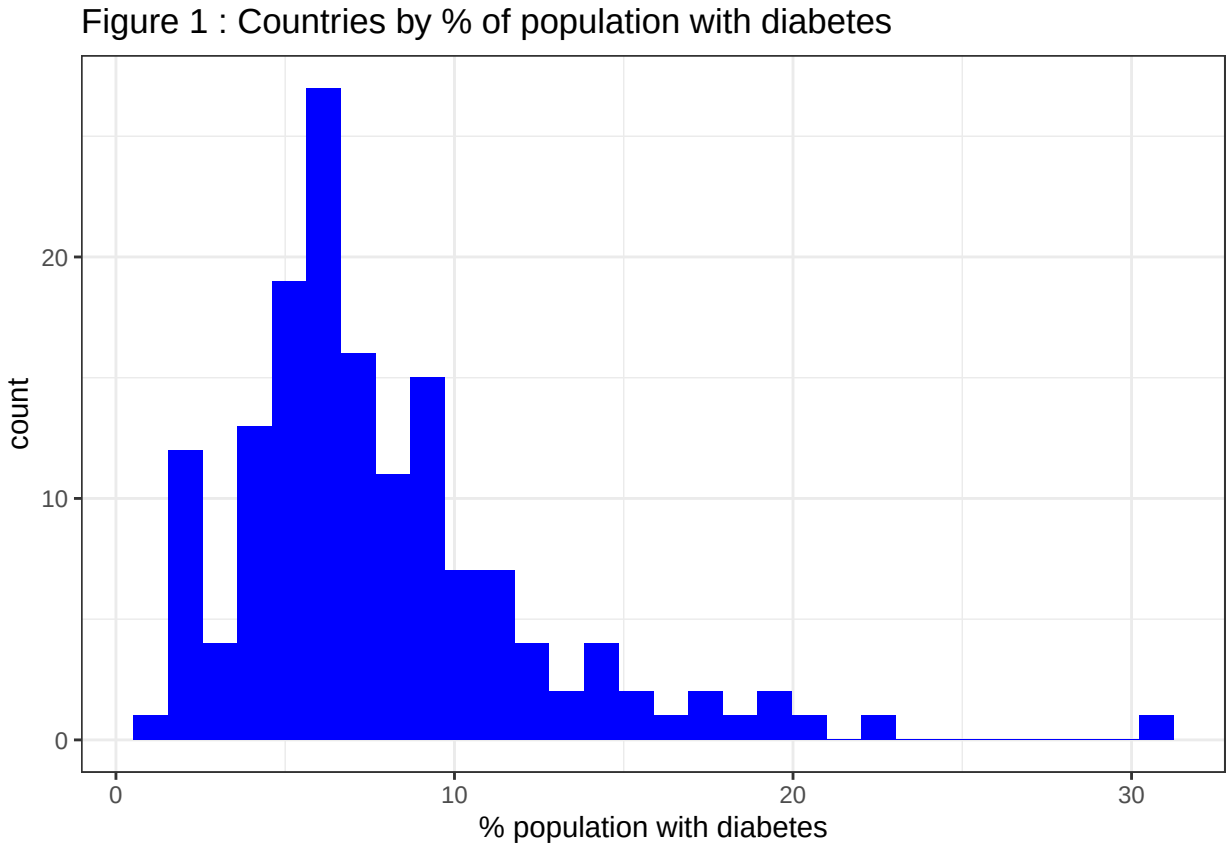

Figure 2: % of population with diabetes and GNI

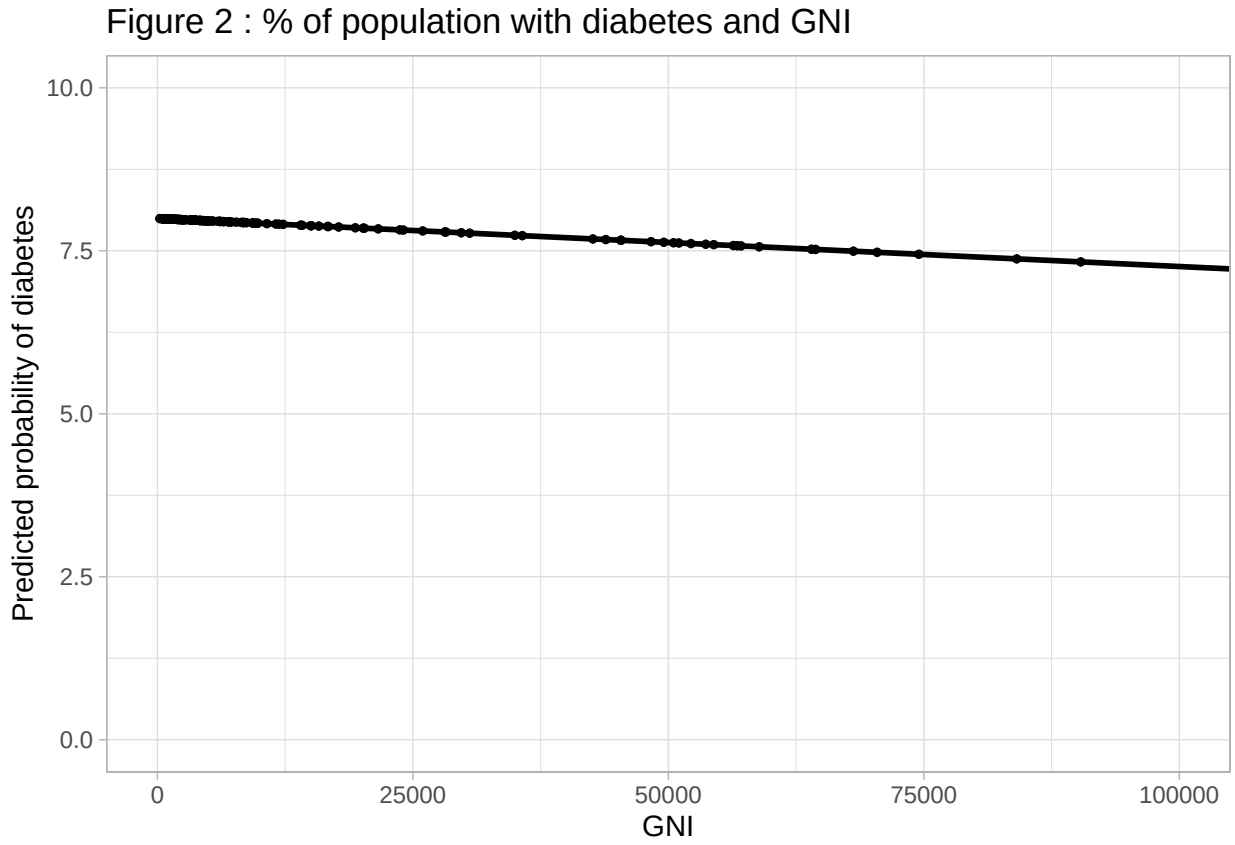

Figure 3: Correlation Matrix