Diabetes Analysis

Sabyasachi Bhoi

2023-02-28

1 Introduction

2 Literature Review

2.1 Paper 1

Non-Communicable Disease Risk Factors and their Trends in India

By Suzanne Nethan, Dhirendra Sinha, and Ravi Mehrotra

About the Paper

In this paper, Nethan, Sinha, and Mehrotra (2017) examine non-communicable diseases (NCD). It accounts for about 38 million (68%) deaths and 5.87 million (60%) worldwide. The paper discusses various national/state-level surveys in India, including single or multiple risk factors. Being overweight, obesity and tobacco use are all covered nationally.

Use Case

As a result of reading this paper, we gained a deeper understanding of what we can expect from a research model as diabetes is a non-communicable disease.

Methodology

Indicator definitions from the World Health Organization (WHO) for both urban and rural populations were used in the research. The percentage of the population covered by the polls for each risk factor was then calculated by adding the state-by-state population proportion and dividing it by the total Indian population. The previous and present data from the periodic surveys were also contrasted to evaluate changes in prevalence. PubMed, Google, and various surveillance systems were searched for this systematic study. Forty-one papers/survey reports from the search results were ultimately determined to be eligible-data on NCD risk variables from states and the nation.

Conclusion

India has a much-delayed response to NCD risk factors surveillance, and information on the exact needs to be more consistent and complete. According to the result, India should plan for a cost and time-effective NCD surveillance system.

2.2 Paper 2

> Performance of Indian Diabetes Risk Score (IDRS) as a screening tool for diabetes in an urban slum

By Lt Col Puja Dudeja, Maj Gurpreet Singh, Maj Tukaram Gadekar, Air Cmde Sandip Mukherji

About the Paper

To identify undetected Type 2 diabetes, the Madras Diabetes Research Foundation (MDRF) created the Indian Diabetes Risk Score (IDRS). In this paper, Dudeja et al. (2017) examine how well the IDRS performs as a screening tool for Type 2 diabetes cases that have not yet been diagnosed and the prevalence of un-diagnosed Type 2 diabetes in urban slums.

Use Case

As a result of reading this paper, we gained a deeper understanding of what we can expect from a research model.

Methodology

Surveys were carried out in urban slum areas. A total of 155 observation IDRS tools, including two adjustable (waist circumference, physical activity) and two non-modifiable (age, family history) risk factors for diabetes, were used to assess the risk of diabetes. Anthropometry data was obtained. Diabetes was diagnosed using fasting venous blood sugar levels.

Conclusion

Using IDRS to the community can effectively detect un-diagnosed diabetes.

2.3 Paper 3

> Prevalence of Diabetes Mellitus and its risk factors

By Akula Sanjeevaiah, Akula Sushmitha, Thota Srikanth

About the Paper

In this paper, Sanjeevaiah, Sushmitha, and Srikanth (2019) explore the prevalence of Diabetes Mellitus (DM) and its risk factors among the population of India. The study found a high prevalence of DM, particularly in urban areas and among those with a family history of the disease. The major risk factors identified were age, obesity, sedentary lifestyle, hypertension, and smoking. The paper highlights the need for early detection and effective management of DM to prevent complications and improve health outcomes.

Methodology

The authors conducted a cross-sectional study for a period of 4 months with a sample size of 250. The basis for the selection of subjects was age greater than 15 years of both genders who are identified with diabetes. Measurements of height and weight were taken to estimate BMI, waist circumference and blood pressure.

Conclusion

This paper concluded that age, waist circumference, hypertension, BMI, smoking habit and total cholesterol are noteworthy when comparing diabetic and non-diabetic subjects.

3 Data Source

It is with utmost pleasure that we elucidate upon the dataset made available by the esteemed National Institute of Diabetes and Digestive and Kidney Diseases, which is designed to predict the likelihood of a patient's susceptibility to diabetes based on specific diagnostic measurements. This dataset is a carefully selected subset of a larger database, wherein all patients are of Pima Indian ancestry and female, aged 21 years or older. The rationale behind this selection criteria is owing to the fact that the Pima Indians exhibit the highest incidence of type 2 diabetes worldwide (Booth et al. 2017).

Our analysis aims to comprehensively examine the most significant factors that contribute to diabetes onset. The dataset comprises nine variables, which were meticulously chosen to ensure their relevance in predicting diabetes. We shall now deliberate on the chosen variables in detail.

3.1 Outcome

The first variable is the dependent variable, "Outcome", which takes binary values of 0 or 1, signifying non-diabetic and diabetic status, respectively.

3.2 Pregnancies

The variable "Pregnancies" records the total number of pregnancies that the patient has experienced. Prior research has established a positive correlation between gestational diabetes mellitus and multiple pregnancies, with women with four or more pregnancies displaying a higher susceptibility to diabetes (Vaajala et al. 2023).

3.3 Glucose

The "Glucose" variable records the plasma glucose concentration (mg/dL) over two hours following an oral glucose tolerance test. High fasting blood glucose concentrations are an indicator of an increased likelihood of developing diabetes. Normal fasting blood glucose concentrations range from 70 mg/dL to 100 mg/dL. Lifestyle changes are recommended if the value lies between 100-125 mg/dL, and diabetes is diagnosed if it exceeds this range.

3.4 Blood Pressure

The "BloodPressure" variable records the average diastolic blood pressure level (mm/Hg) of the patient. Patients with hypertension exhibit insulin resistance and are at a higher risk of developing diabetes. A staggering 66% of diabetic individuals worldwide have blood pressure levels greater than 130/80 mm Hg or are diagnosed with hypertension (Petrie, Guzik, and Touyz 2018).

3.5 Skin Thickness

The "SkinThickness" variable records the triceps skin fold thickness (mm) of the patient. In type-2 diabetes, insulin resistance affects subcutaneous tissue thickness. Previous research has indicated an association between diabetic patients and increased skin thickness (Jain et al. 2013).

3.6 Insulin

The "Insulin" variable records the result of the 2-hour serum insulin (mu U/ml) level. Healthy individuals exhibit insulin levels between 5 and 15 µU/mL, whereas higher insulin levels are directly correlated with diabetes.

3.7 BMI

The "BMI" variable records the body mass index of the patient. Overnutrition and obesity may result in insulin resistance, leading to increased glucose levels in the blood and eventually, diabetes. Research has shown that individuals with a BMI greater than 25 are more likely to develop diabetes.

3.8 Diabetes Pedigree Function

The “DiabetesPedigreeFunction” variable measures the likelihood of becoming diabetic based on family history. A family tree is generated based on which a function is generated, assigning a high or low value based on the number of diabetic patients in the patient’s family history (Massaro et al. 2022).

3.9 Age

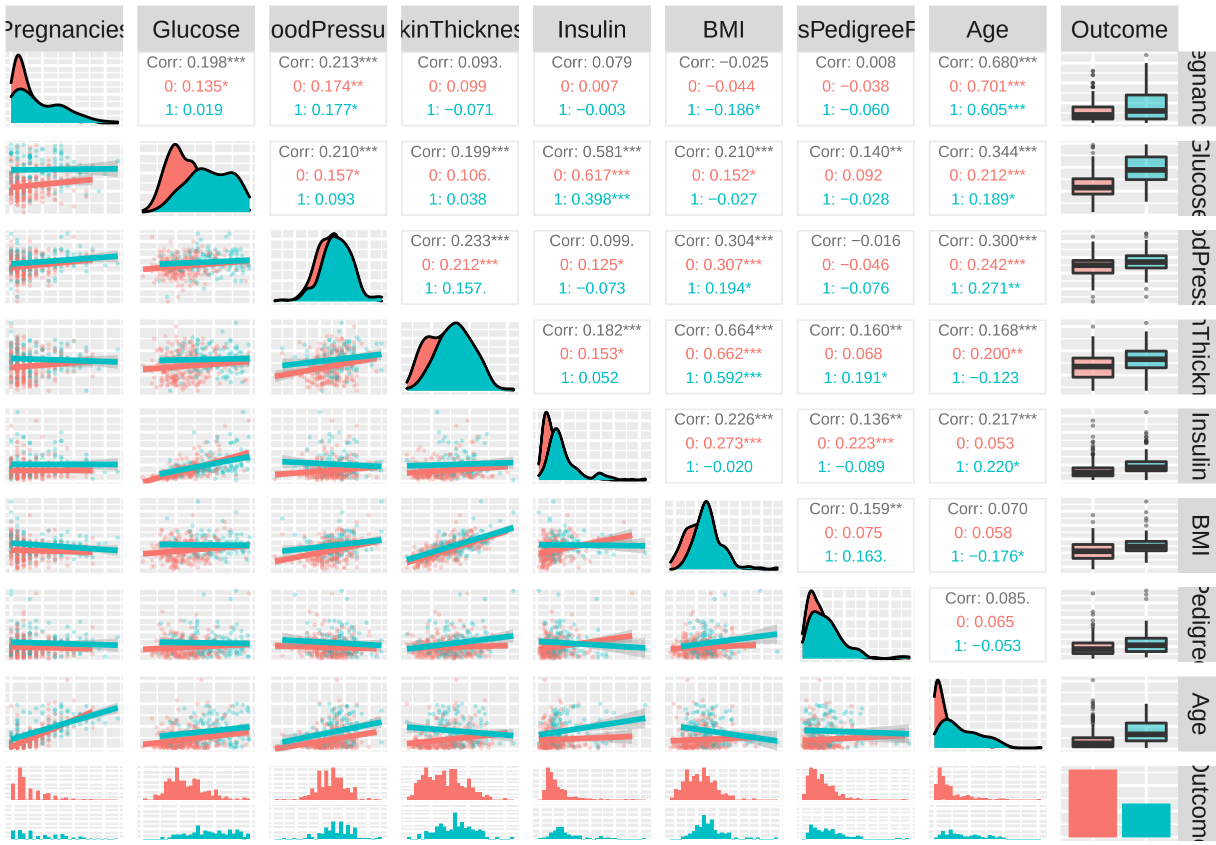
Finally, the “Age” variable records the age of the patient. With age, individuals are more likely to develop multiple medical conditions, including high blood pressure and high cholesterol. Current literature suggests that middle-aged individuals are more likely to develop diabetes due to these indirect factors.

In conclusion, this dataset is a valuable resource for predicting diabetes onset based on specific diagnostic measurements. Our analysis of the chosen variables underscores the significance of hypertension, insulin resistance, and family history in determining an individual’s susceptibility to diabetes.

4 Summary Statistics

Table 1: Summary statistics of the dataset

Statistic	N	Mean	St. Dev.	Min	Max
Pregnancies	392	3.301	3.211	0	17
Glucose	392	122.628	30.861	56	198
BloodPressure	392	70.663	12.496	24	110
SkinThickness	392	29.145	10.516	7	63
Insulin	392	156.056	118.842	14	846
BMI	392	33.086	7.028	18.200	67.100
DiabetesPedigreeFunction	392	0.523	0.345	0.085	2.420
Age	392	30.865	10.201	21	81



Pregnancies
Glucose
oodPressu
kinThicknes
Insulin
BMI
sPedigreeF
Age
Outcome
Corr: 0.198***
0: 0.135*
1: 0.019
Corr: 0.213***
0: 0.174**
1: 0.177*
Corr: 0.093.
0: 0.099
1: −0.071
Corr: 0.079
0: 0.007
1: −0.003
Corr: −0.025
0: −0.044
1: −0.186*
Corr: 0.008
0: −0.038
1: −0.060
Corr: 0.680***
0: 0.701***
1: 0.605***
Corr: 0.210***
0: 0.157*
1: 0.093
Corr: 0.199***
0: 0.106.
1: 0.038
Corr: 0.581***
0: 0.617***
1: 0.398***
Corr: 0.210***
0: 0.152*
1: −0.027
Corr: 0.140**
0: 0.092
1: −0.028
Corr: 0.344***
0: 0.212***
1: 0.189*
Corr: 0.233***
0: 0.212***
1: 0.157.
Corr: 0.099.
0: 0.125*
1: −0.073
Corr: 0.304***
0: 0.307***
1: 0.194*
Corr: −0.016
0: −0.046
1: −0.076
Corr: 0.300***
0: 0.242***
1: 0.271**
Corr: 0.182***
0: 0.153*
1: 0.052
Corr: 0.664***
0: 0.662***
1: 0.592***
Corr: 0.160**
0: 0.068
1: 0.191*
Corr: 0.168***
0: 0.200**
1: −0.123
Corr: 0.226***
0: 0.273***
1: −0.020
Corr: 0.136**
0: 0.223***
1: −0.089
Corr: 0.217***
0: 0.053
1: 0.220*
Corr: 0.159**
0: 0.075
1: 0.163.
Corr: 0.070
0: 0.058
1: −0.176*
Corr: 0.085.
0: 0.065
1: −0.053
egnanc
Glucose
odPress
Thickn
Insulin
BMI
Pedigre
Age
utcom

5 Regression

Table 2: Robust Regression Results

	Dependent variable:
	Outcome
Pregnancies	0.040
	(0.026)
SkinThickness	0.001
	(0.003)
Glucose	0.008***
	(0.001)
BloodPressure	0.001
	(0.002)
DiabetesPedigreeFunction	0.203***
	(0.059)
Age	−0.001
	(0.005)
BMI	0.010**
	(0.004)
Insulin	−0.002***
	(0.001)
Pregnancies:Age	−0.001
	(0.001)
Age:Insulin	0.00004***
	(0.00001)
Constant	−1.140***
	(0.178)
Observations	392
Residual Std. Error	0.344 (df = 381)
Note:	*p<0.1; **p<0.05; ***p<0.01

5.1 Diagnostics

Test for Heteroskedasticity

```
##
##  studentized Breusch-Pagan test
##
## data:  fit
## BP = 59.857, df = 10, p-value = 3.856e-09
```

Test for Omitted Variable Bias

We're going to employ the Ramsey RESET Test (Ramsey 1969) to test for Omitted Variable Bias.

```
##
##  RESET test
##
## data:  fit
## RESET = 1.0083, df1 = 1, df2 = 380, p-value = 0.3159
```

6 References

Booth, Clayton, Maziar M Nourian, Shannon Weaver, Bethany Gull, and Akiko Kamimura. 2017. "Policy and Social Factors Influencing Diabetes Among Pima Indians in Arizona, USA." *Policy* 7 (3).

Dudeja, Puja, Gurpreet Singh, Tukaram Gadekar, and Sandip Mukherji. 2017. "Performance of Indian Diabetes Risk Score (IDRS) as Screening Tool for Diabetes in an Urban Slum." *Medical Journal Armed Forces India* 73 (2): 123–28. https://doi.org/https://doi.org/10.1016/j.mjafi.2016.08.007.

Jain, Sunil M, Kirnesh Pandey, Alok Lahoti, and P Kiran Rao. 2013. "Evaluation of Skin and Subcutaneous Tissue Thickness at Insulin Injection Sites in Indian, Insulin Naive, Type-2 Diabetic Adult Population." *Indian Journal of Endocrinology and Metabolism* 17 (5): 864–70.

Massaro, Alessandro, Nicola Magaletti, Gabriele Cosoli, Vito Giardinelli, and Angelo Leogrande. 2022. "The Prediction of Diabetes." *Available at SSRN 4135264*.

Nethan, Suzanne, Dhirendra Sinha, and Ravi Mehrotra. 2017. "Non Communicable Disease Risk Factors and Their Trends in India." *Asian Pacific Journal of Cancer Prevention: APJCP* 18 (7): 2005.

Petrie, John R, Tomasz J Guzik, and Rhian M Touyz. 2018. "Diabetes, Hypertension, and Cardiovascular Disease: Clinical Insights and Vascular Mechanisms." *Canadian Journal of Cardiology* 34 (5): 575–84.

Ramsey, J. B. 1969. "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis." *Journal of the Royal Statistical Society. Series B (Methodological)* 31 (2): 350–71. http://www.jstor.org/stable/2984219.

Sanjeevaiah, Akula, Akula Sushmitha, and Thota Srikanth. 2019. "Prevalence of Diabetes Mellitus and Its Risk Factors." *IAIM* 6 (3): 319–24.

Vaajala, Matias, Rasmus Liukkonen, Ville Ponkilainen, Maiju Kekki, Ville M Mattila, and Ilari Kuitunen. 2023. "Higher Odds of Gestational Diabetes Among Women with Multiple Pregnancies: A Nationwide Register-Based Cohort Study in Finland." *Acta Diabetologica* 60 (1): 127–30.