Introduction

The goal of the project was to find out whether survey questions from the collected data could provide accurate predictions of diabetes, find out the most predictive risk factors of illness to help local authorities in developing and implementing strategies to identify risk groups and tackle risk factors of diabetes.

Analysis

We used four datasets containing information about 253 680 individuals. 22 metrics for each individual were considered, including socio-demographic, behavioral factors and medical history .

We decided to focus on the biosocial and behavioral factors which had the greatest impact on diabetes and the risk factors which could lead to diabetes: BMI, physical activity, income and education level.

Conclusion

14% of individuals in the dataset are diabetic.

The data shows that incidences of diabetes in every five year age range increase by 2-3% from 1% in age range 18-24 to 9% in age range 45-49 reaching its maximum of 22% in age range 70-74.

In each age range up to 40-44 the incidence of diabetes in females is 1% more than that in males. At around 40 years, the proportion of diabetic males exceeds the proportion of diabetic females. The difference in diabetes in the two groups increases from 1% in 40-45 to 2% up to age 65 then increasing to 4%-6%.

The data shows that from age 18 to around 40, the obesity population increases from 22% to 36%. The amount of obesity then levels off staying at 37%-38% until age 70 when it falls to the level of 30-34 years old - 34% before decreasing to 20% during the later stages of life. The data for overweight people follows a similar trend.

Known factors contributing to obesity include eating habits and physical activity. We have found a negative correlation between the amount of fruits and vegetables eaten, physical activity and BMI.

- 22% of obese people and 18% of overweight people were found not to eat vegetables versus 16% of people with a normal BMI.
- 42% of obese and 36% of overweight people were found not to eat fruit whereas 31% of people with a normal BMI were found not to eat fruit.
- 22% of obese people and 18% of overweight people don't exercise enough compared to 16% of people with a normal BMI.

Underweight people do not appear to follow this trend. 35% don't eat fruit, 20% don't eat vegetables and 20% don't exercise. However, underweight people made up only 1% of the dataset.

We tried to understand the reasons for these trends using available metrics in the dataset and found that there is a positive correlation within all the factors of level of education, income range and fruit/vegetable consumption. Physical activity as a factor follows a similar trend.

- 54% of individuals with income <15K are eating enough fruit against 66% with income >50K.
- 69% of individuals with income <15K are eating enough vegetables against 86% with income >50K.
- 60% of individuals with income <15K exercise enough versus 83% of people with income >50K.
- 54% of individuals who didn't graduate high school are eating enough fruit against 69% with college or similar education.
- 68% of individuals who didn't graduate high school are eating enough vegetables against 87% with college or similar education.
- 57% of individuals who didn't graduate high school do exercise enough versus 85% with college or similar education.

Unsurprisingly, there is a positive correlation between level of education and income range.

As a result of these trends 25% of individuals with income <15K have diabetes versus 9% diabetic individuals with income > 50K. There is a negative correlation between the occurrence of diabetes in individuals and the amount they earn.

26% of individuals who didn't graduate high school have diabetes versus 10% diabetic individuals who graduated college or similar. There is a negative correlation between the amount of individuals who have diabetes and their education level.

Although generally there is a positive correlation between the occurrence of diabetes in individuals and BMI across all income groups, in the lower income groups the incidence of diabetes is often seen even amongst those who are underweight or have a normal BMI. In individuals with higher income, the occurrences of diabetes are seen mainly in overweight and obese people.

In a similar way to income, the lower the education level, the greater the incidence of diabetes amongst all individuals. However, where the education level is lower, diabetes is also seen in individuals who are underweight or who have a normal BMI. In individuals with higher education levels, the occurrences of diabetes are seen mainly in overweight and obese people.

Recommendations

From analysis of this dataset, it is possible to suggest that maintaining adequate nutrition and physical activity could decrease the likelihood of obesity and therefore diabetes.

Individuals with a lower income and/or education level eat fewer fruit and vegetables and are less physically active. This can then lead to diabetes.

Age appears to be a significant factor in the onset of type 2 diabetes. Males also seem to be at higher risk after around 40 years.

Since it's difficult to influence socio-demographic factors we would recommend increasing campaigns that promote a healthy lifestyle, provide guidance on the risk factors involved in diabetes and highlight the importance of healthy eating and physical activity. This would be relevant to individuals across all income ranges but special attention should be given to those on lower incomes.

To increase the benefits of these campaigns, they should be started as early as possible, ideally at primary school level.

There is also a need to raise local awareness of·the risk factors associated with Type 2 diabetes, the importance of seeking a risk assessment and the benefits of early diagnosis.

Data Enrichment

The data in this project omits relevant characteristic information for the individuals such as race/ethnicity and family history of diabetes. For a more thorough and detailed analysis, additional data would need to be provided.