STEM Fellowship Big Data Challenge 2020

Exploring the Correlation Between Depression and Obesity Across the World

Daniel Tsai[1] and Linda Nong[2]

[1]University of Toronto
[2]University of Guelph

June 3, 2020

Abstract

Depression and obesity are widespread illnesses that are increasing in prevalence across the world. Suicide due to depression is a leading cause of death for youth and obesity appreciably increases the risk of other life-threatening medical conditions. This study aims to understand the correlation between these diseases on a global scale. The average BMI of a country was used as an indicator for the prevalence of obesity and depression levels were estimated as a percentage of the total population. Secondary data from 2015 was collected from the World Health Organization and a regression analysis was performed to determine the relationship between the prevalence of depression and the average BMI of a country. The results indicated a moderate, positive linear relationship. This was further supported by a low p-value (1.38e-8) and a categorical analysis which determined that, on average, countries with higher depression percentages had a higher mean BMI. In terms of modeling this relationship, a low R-squared value (16.5%) suggested that the prevalence of depression cannot be directly used to model and predict the mean BMI for a given country. Future statistical models on this topic should include common risk factors to both obesity and depression, which include lifestyle habits, genetics, and social environment. An accurate model can help health officials identify populations at risk for developing high levels of depression or obesity based on demographic information, and target specific risk factors to decrease that risk.

Keywords
Depression, Obesity, BMI, Mental Health, Statistical Modeling

1 Introduction

Mental health disorders exist everywhere across the globe, and depression is one of the most common. Suicide due to depression accounts for nearly 20 % of adolescent deaths in the United States in 2017 [1] [2]. Furthermore, depression rates have been increasing over the last few decades [3]. This increase has be attributed to a wide range of factors, including unchanged mental health treatments, a sedentary lifestyle, and loneliness due to the usage of social media platforms [3] [4]. Depression involves depressive episodes of varying severity, and is characterized by a pessimistic mood, loss of interest and enjoyment, reduced self-confidence, and reduced energy [5]. As well, depression is frequently associated with other psychological and physiological illnesses, and the development of physical diseases increases the risk for depressive illnesses [6].

Obesity is another common disease across the world; its prevalence has nearly tripled since 1975 and affects more than 13% of the world's adult population. It is associated with an increased risk for medical conditions such as cardiovascular disease, musculoskeletal disorders, diabetes, and certain cancers, which can be life-threatening [7]. Body Mass Index (BMI) is a simple method used to provide a rough assessment of someone's body condition, using the ratio of their height and weight. An individual is considered overweight when their BMI is greater than 25, and considered obese when their BMI is greater than 30 [7]. Fundamentally, obesity results from a long-term positive energy balance in the body, due to an increase in the intake of high-energy foods and a decrease in physical activity. This means that obesity is more easily preventable, compared to psychological disorders like depression. However, obesity is also affected by multitudes of factors, such as environment, stress, genetics, sleep habits, and medications [8].

Both obesity and depression are widespread, potentially life-threatening diseases that are affecting more people every year. There is an interest in the possibility of depression increasing the risk for obesity, and vice versa, since both illnesses seem to have multitudes of risk factors. Both depression and obesity share the common risk factors of having a sedentary lifestyle, cer-

tain genetics, excessive stress, and negative social environment. Many studies have shown a significant positive correlation between depression and obesity in individuals, and found a bidirectional association between the two diseases [9] [10] [11]. However, those studies only looked at the prevalence of depression and obesity within a small sample or experimental population. Understanding the relationship between depression and obesity on a larger, population-level scale rather than individual basis will show a broader perspective and provide insight on how to best combat these illnesses.

This study aims to explore this correlation on a global scale by analyzing the average BMI and frequency of depression in different countries. A linear model was also created to predict the average BMI given the frequency of depression of a country. In order to do this, datasets were extracted from the World Health Organization (WHO) website, describing the frequency of depression in a country and the mean BMI of a country. A regression analysis was performed on the data to determine a correlation, and the p-value, pearson coefficient, and R-squared value were also calculated. A positive relationship between obesity and depression would indicate that countries with higher prevalence of obesity also have a higher prevalence of depression. In light of decreasing the number of individuals with obesity and depression, a positive correlation would indicate that targeting the reduction of one disease could potentially indirectly reduce the prevalence of the other disease.

2 Materials & Methods

The datasets (processed CSV tables) for this study were taken from the World Health Organization's (WHO) Global Health Observatory because of the organization's reliability for information and their advocacy for improving global public health. As an international organization, the WHO has the most extensive data collection on statistics from countries across the world, which is the subject of interest. Since two different datasets were used to test the hypothesis, having both datasets generated by the same organization will also help with data merging because of the similarity of the table formats. The dataset for depression ("Population based prevalence of depression")[12] describes the estimated percentage of the population of the sampled countries that were diagnosed with depression in 2015. For an individual to be classified as depressed, they would have symptoms such as loss of interest and enjoyment, and decreased energy [13]. The dataset for mean BMI ("Mean BMI among adults")[14] displays the total mean BMI (Body Mass Index) for the studied countries and the mean BMI for both male and females, separately. The BMI values in the BMI dataset were calculated using the height and weight of the participants from samples of the population.

The CSV files were imported onto a Jupyter Notebook for data analysis, using the Pandas library. In this study, mean BMI was used as an indicator for the frequency of obesity in a country. For the data cleaning process, the BMI dataset was narrowed down to four columns: the country of study, mean BMI of both sexes, and separate mean BMIs of males and females for the year 2015. The data included a confidence interval for each mean BMI value, which was removed. After merging the mean BMI dataset with the depression dataset, it was noted that certain countries had only one of mean BMI or estimated depression values, so those countries were dropped from the study. Otherwise, the missing values would have to be estimated with the overall average and result in inaccurate data. In total, 12 countries were removed from the study because they were missing either the depression percentage or the mean BMI values, and the final analysis was performed on the remaining 181 countries.

The first statistical analysis performed was to determine the pearson coefficient and the p value of the correlation between the datasets. This was calculated using the pearsonr function (imported from the scipy library). Data visualizations including a regression plot and residual plot were generated through Python's Seaborn library. The prediction model was developed using the sklearn library's LinearRegression to create a linear model. In order to create a more accurate prediction, it was determined that a polynomial fit model should be used. The r2_score and mean_squared error functions (imported from the sklearn library) were used to calculate the R-squared and mean squared error values for the linear regression and polynomial fit models.

To find the differences between countries with overall higher, medium, and lower depression percentages, the depression dataset was split into three groups using the numpy library's linspace function. Three equal sized bins were created, and the category classifications and average BMI for each bin were added into the dataset as new columns.

3 Results

The p-value of the regression plot (Figure 1) for the relationship between depression percentage and mean BMI is 1.38e-8. Each dot on the plot represents a country from the dataset and the horizontal and vertical axes represent the prevalence of depression and total mean BMI (both male and female) of the population of each country, respectively. The regression plot aims to minimize the total distance between all the observed values (the dots) and the fitted value (the regression line). The regression line in the plot in Figure 1 has a positive slope, meaning that as the prevalence of depression increases, the mean BMI increases as well. This plot gives an initial indication that there is a positive linear relationship between depression levels and average BMI.

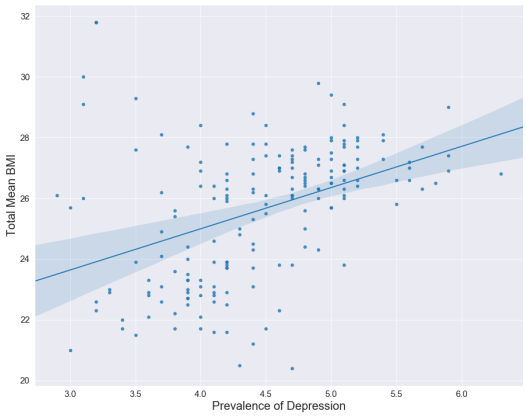

Figure 1: A Regression Plot for the Mean BMI of 181 Countries as a Function of the Prevalence of Depression

The pearson coefficient, which measures the strength and direction of a statistical relationship, falls at around 0.406. The model created in this study for the prevalence of depression and average BMI can be better predicted by using a polynomial fit instead of linear regression, since a higher R-squared value was calculated with the polynomial model (27.00%) than the linear model (16.51%).

The dataset for estimated depression has a range of 2.9% to 6.3%, with 31 unique values. 3 equal categorical bins (low, medium, high) were created for these values, in order to find the average BMI differences between countries with low, medium, and high depression percentages. This was done using the numPy library's linspace() function to yield the following values:

Depression Bins	Avg. BMI (Both Sexes)	Avg. BMI (Male)	Avg. BMI (Female)
Low	24.51	23.84	25.17
Medium	25.81	25.47	26.17
High	27.18	27.07	27.29

Table 1: Classification of depression percentages into Low, Medium and High with Average BMIs for each category

The values in Table 1 show that the average BMI is lowest (24.51) for the low depression category and highest (27.18) for the high depression category, which fortifies the positive correlation between the two variables.

4 Discussion

The low p-value of 1.38e-8 means there is strong evidence that the correlation between the depression percentage and the mean BMI is statistically significant. In addition, the regression plot in Figure 1 shows that across the countries studied, as the prevalence of depression increases, the mean BMI also increases. Therefore, these datasets provide strong evidence to reject the null hypothesis that depression does not have any correlation with BMI.

A pearson coefficient of 0.406 represents a moderately weak, positive relationship between the two variables. At this point of the analysis, the moderate coefficient indicates that the percentage of depression within a population can potentially be a good predictor of the average BMI of a country. The positive correlation indicates that an increase in depression prevalence predicts an increase in average BMI.

To develop a prediction model, Linear Regression was used first. The R-squared value and mean squared value were determined, but the results showed a low R-squared value (16.51%) and a relatively high mean squared value (4.25). The next model to try was the polynomial fit model. Results were slightly better, but still yielded a low R-squared value (29.35%) and a high mean squared value (3.60). This indicates that this model only accounts for 29.35% of all the variability of the data around its mean. Even with the polynomial fit model, the R-squared value is too low to confidently predict the value of the mean BMI given the depression percentage of surveyed countries. This suggests that the independent variable (prevalence of depression) cannot be used to accurately predict the dependent variable (mean BMI), even though the low p-value indicates that the correlation between depression and mean BMI is statistically significant.

From the categorical bins created, it is evident that a depression prevalence that is classified as "High" has the highest average BMI, while "Low" depression prevalence has the lowest average BMI. This indicates that a country with a higher percentage of depression within the population tends to have a higher average BMI. There are many reasons that could cause this correlation, such as a poorly funded health care system resulting in higher levels of all illnesses. Additionally, there are several risk factors shared by obesity and depression, including poor lifestyle habits and social environment, and it is possible that certain countries have high levels of both depression and obesity because of the status of these risk factors.

Additionally, a meta-analysis done in 2018 has determined that there are several biological pathways shared by obesity and depression, including inflammation, HPA dysregulation, and leptin/insulin dysregulation [15]. Since these pathways are shared by obesity and depression, a dysfunction can trigger the development of both diseases. Another study showed that depressed individuals are at a significantly higher risk of developing obesity compared to undepressed individuals [16], which would explain the positive correlation between depression and mean BMI on a population level, as seen in this exploration. A study has

even proposed that depression and obesity are separate expressions of the same disease [17]. It has been determined that the treatment of one of obesity or depression appears to improve the other condition as well [18]. This is valuable information for improving the health of a global population. By increasing the accessibility of social health support to reduce the prevalence of depression, obesity can be indirectly reduced as well. Similarly, encouraging a healthy lifestyle to reduce obesity can indirectly reduce depression. Furthermore, since these two diseases are so closely related, an interesting topic to look into is whether or not medication and treatments targeted at depression are effective at targeting obesity, and vice versa.

There are several things that must be kept in mind regarding the data that was used in this study. The WHO is a globally trusted entity for its reliable data collection from countries all over the world. However, majority of the depression-related data was estimated through statistical modeling, using medical data and surveys of a country. This is because raw data is scarce in lower-income countries, especially regarding mental health, which is not high priority for those countries. Thus, it is important to acknowledge that there are varying levels of accuracy for the percentages of depression in different countries, depending on the statistical modeling used. Additionally, mental illnesses, notably depression, often go undiagnosed in less severe cases. Depending on available health services, the estimated values of the prevalence of depression could be lower than the actual values.

A limitation of this study is that only one year of data was analyzed, rather than an ongoing trend of data year after year. Further analysis on data from a longer time period is needed to see whether changes in depression percentages actually have a direct affect on the mean BMI of each country. It is highly likely that an increase in depression rates does not directly cause an increase in obesity rates, as there are many other factors contributing to an increase in BMI. For the world-wide range of this study, these factors include the geography of a country and the level of development of a country, which influence the food availability in that region.

Furthermore, using secondary sources of data is always limited by the uncertainty of the methods of collection for that dataset. This paper is not able to comment on the accuracy of the data used or the methods used to sample and model the various populations.

Conclusions

In this investigation, it was determined that there is a positive correlation between the prevalence of depression and the mean BMI (as an indicator of the prevalence of obesity) in countries across the world. However, the correlation is not strong enough to accurately predict the mean BMI of a country based solely on the prevalence of depression in that country. There are several common factors that affect the risk of developing depression and obesity, including lifestyle habits (ie sleeping, eating, and physical exercise), genetics, and social environment. Further research can be done to create a model, using these risk factors, that more accurately predicts the risk for a population to develop higher levels of obesity based on their prevalence of depression, and vice versa. Based on the model results, health officials can better target certain factors to reduce the prevalence of the two illnesses. Additionally, the strong association between depression and obesity raises the question of whether or not current medication for depression could be effective for the treatment of obesity.

Acknowledgements

Thank you to STEM Fellowship for providing this opportunity to engage in data-driven research, and for providing resources for this Big Data Challenge. As well, thank you to Mohammed Khalid for his support and advice throughout the development of this study.

References

[1] Melonie Heron. Deaths: Leading causes for 2017, Jun 2019.

[2] Hannah Ritchie and Max Roser. Mental health. *Our World in Data*, 2018. https://ourworldindata.org/mental-health.

[3] Ramin Mojtabai, Mark Olfson, and Beth Han. National trends in the prevalence and treatment of depression in adolescents and young adults. *Pediatrics*, 2016.

[4] Sungkyu Park, Inyeop Kim, Sang Won Lee, Jaehyun Yoo, Bumseok Jeong, and Meeyoung Cha. Manifestation of depression and loneliness on social networks: A case study of young adults on facebook. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work Social Computing*, CSCW '15, page 557–570, New York, NY, USA, 2015. Association for Computing Machinery.

[5] Clinical descriptions and diagnostic guidelines.

[6] Guy M Goodwin. Depression and associated physical diseases and symptoms, Jun 2006.

[7] Obesity and overweight, Apr 2020.

[8] Obesity, Feb 2020.

[9] Leonore [de Wit], Floriana Luppino, Annemieke [van Straten], Brenda Penninx, Frans Zitman, and

Pim Cuijpers. Depression and obesity: A meta-analysis of community-based studies. *Psychiatry Research*, 178(2):230 – 235, 2010.

[10] Albert J Stunkard, Myles S Faith, and Kelly C Allison. Depression and obesity. *Biological Psychiatry*, 54(3):330 – 337, 2003.

[11] Gregor Hasler, Shmuel Lissek, Vladeta Ajdacic, Gabriella Milos, Alex Gamma, Dominique Eich, Wulf Rössler, and Jules Angst. Major depression predicts an increase in long-term body weight variability in young adults. *Obesity Research*, 13(11):1991–1998, 2005.

[12] Gho | by category | risk factors - population-based prevalence of depression, Feb 2018.

[13] Indicator metadata registry details.

[14] Gho | by category | mean body mass index trends among adults, crude (kg/m^2) - estimates by country, Sept 2017.

[15] Yuri Milaneschi, W. Kyle Simmons, Elisabeth F. C. Van Rossum, and Brenda Wjh Penninx. Depression and obesity: evidence of shared biological mechanisms. *Molecular Psychiatry*, 24(1):18–33, 2018.

[16] Bruce Blaine. Does depression cause obesity? *Journal of Health Psychology*, 13(8):1190–1197, Nov 2008.

[17] Roland Rosmond. Obesity and depression: same disease, different names? *Medical Hypotheses*, 62(6):976 – 979, 2004.

[18] Nattinee Jantaratnotai, Kanokwan Mosikanon, Yena Lee, and Roger S. McIntyre. The interface of depression and obesity. *Obesity Research Clinical Practice*, 11(1):1 – 10, 2017.