Causal Inference with Mendelian Randomization to Explore Risk Factors of Diabetes

Amulya Agrawal[1]

[1]Jefferson City High School, Jefferson, MO, USA

ABSTRACT

Since the dawn of time, health conditions have dictated life around the world. Gradually, through the advancement of medicine and technology, more and more treatments have been discovered to combat these conditions. One of these conditions is a disease known as diabetes. Even with a plethora of treatments being utilized by individuals internationally, diabetes continues to be the one of the leading causes of death worldwide [1]. Caused by a deficiency of insulin, a hormone created and released by the pancreas, diabetes renders individuals unable to effectively utilize glucose. With low amounts of insulin, cells cannot allow glucose to enter them and be used as energy for the body, leaving high amounts of glucose to build up in the bloodstream. Several factors have been suggested as a link to causing this insulin deficiency, resulting in diabetes. However, it is important to remember that there are two types of diabetes present, Type 1 and Type 2 diabetes. This report uses Mendelian Randomization to analyze contributing factors of Type 1 and 2 diabetes and explain the roles of confounds and genetics in the pathogenesis of the disease.

Introduction

Diabetes is the one of the leading causes of death for millions internationally [1]. This makes it imperative to identify and analyze contributing factors in-order to prevent more from developing the disease. Prevention is especially important because, as of now, there are no treatments to cure diabetes [2]. While insulin treatments are available, it is only a temporary palliative, not a permanent cure. Insulin dosages must be taken frequently by individuals to help cells absorb glucose, which is necessary for the glycolysis process.

Decreased insulin production leads to high amounts of glucose remaining in the bloodstream, which causes a condition known as hyperglycemia. This results in osmotic imbalances in the bloodstream, which causes symptoms such as high thirst levels, excessive urination, significant changes in weight, dry skin, troubles in wound healing, and high hunger levels [3]. Knowing this, one can apply these ideas and concepts to distinguish between the two types of diabetes – Type 1 and Type 2 diabetes.

The Etiology of Diabetes

Type 1 diabetes is most commonly known to occur in younger people. Hence, it is often referred to as "juvenile diabetes." It is also known as Insulin Dependent Diabetes Mellitus (IDDM) and is the least common type of diabetes to be diagnosed [4]. Type 1 diabetes specifically occurs when there is a genetic mutation that results in the body destroying pancreatic cells, which are responsible for producing insulin. Type 2 diabetes, on the other hand, is the more commonly diagnosed type of diabetes, which is mostly seen in middle-aged and older people [5]. It is also known as Non-Insulin Dependent Diabetes Mellitus (NIDDM). Type 2 diabetes specifically occurs when pancreatic cells do not produce insulin in sufficient amounts or do not utilize the insulin production properly. A unique aspect about Type 2 diabetes is that, for many, it can be prevented. Most diagnosed with Type 2 diabetes exhibit underlying

physical causes such as being overweight or living a largely sedentary lifestyle [6]. This, in turn, results in Type 2 diabetes to be regarded as "progressive" and having the capability to develop into more harmful health scares, such as cardiovascular disease.
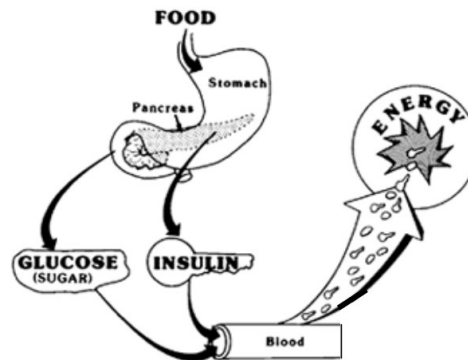

Fig. 1: [7] This diagram displays the processes secreted insulin goes through. Insulin, which is produced by the metabolism of food, has a function of absorbing glucose from the bloodstream. Cells can then utilize the glucose to create energy for body processes. However, when one is diagnosed with diabetes, this process does not occur as smoothly. In Type 1 diabetes, insulin is not able to be produced due to pancreatic cells being destroyed. In Type 2 diabetes, insulin produced by pancreatic cells are not utilized properly, and/or sufficient amounts of insulin are not produced.

Casual Inference with Mendelian Randomization

To analyze contributing factors of both types of diabetes, it is important to first understand the role of genetics and then how Mendelian Randomization and confounds play a role in development. While genetics is more commonly known to play a role in Type 1 diabetes, it can also play a partial role in the development of Type 2 diabetes. For instance, if one's family members have diabetes, then the individual would have a higher chance of contracting diabetes. If it is Type 2 diabetes, then actions such as exercising and maintaining a healthy diet can reduce the impacts of the development of the disease.

Mendelian Randomization would help with identifying different exposure factors and measuring their effects, which in this case would entail evaluating the development and presence of diabetes. Specific exposure factors evaluated in this paper include Body Mass Index (BMI), High Blood Pressure Levels, and Smoking. The outcome, specifically, is whether or not Type 1 or Type 2 diabetes occurs. Mendelian Randomization utilizes genetic variables to identify risk factors that affect the development of the disease. Confounds would then help analyze external factors that may affect the development of diabetes, but they are not directly involved or acknowledged as a primary exposure variable.

The challenge that needed to be dealt with in this research was being able to differentiate between correlation and causation. Specifically, as mentioned previously, confounds could alter the validity of results. If a risk factor and outcome appear together at the same time, it would be difficult to distinguish whether the risk factor is causing the outcome, vice versa, or if there's another factor relating to the outcome and the risk factor. Hence, Mendelian Randomization was utilized to identify causal relationships between risk factor and outcome.

Statistical Analysis with R Programming

To approach this, R programming was utilized to evaluate and build off of the research of other researchers. Statistics were collected to deny or accept certain claims and factors. More specifically, an MRBase database, which is

interfaced with R, was utilized to collect data to find correlations and relationships between chosen exposures and the outcomes of both Type 1 and 2 diabetes.

Methods

Mendelian Randomization

The usage of Mendelian Randomization to determine relationships between chosen factors and outcomes is the best approach because in Mendelian Randomization analyses, variants can act as exposures in an environment independent of confounders. Confounds can result in sways in data, causing inaccurate results, which can be avoided through this method of analysis.
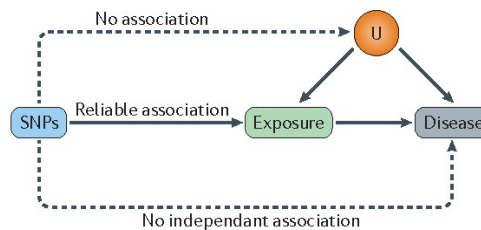

Fig. 2: [8] Mendelian Randomization identifies a variety of different exposure factors of certain outcomes by investigating its causal relationships. In an ideal situation, the outcome would only be influenced through the exposure. However, confounds (U) can affect outcomes, so it is imperative to not associate external factors with the outcome. The single nucleotide polymorphism (SNP) must only result in the disease outcome if it includes the specific exposure being chosen for.

MRBase Database

The usage of an MRBase database in this study specifically gathers data from studies of white European populations of a variety of ages. Unfortunately, the database did not offer a significant amount of diversity, as most studies focused largely on this demographic. The MRBase database was interfaced with R, a programming language that helps statisticians and researchers find data to conduct data analysis in a variety of areas. All of the data collected dealt with exposure to certain factors and calculating the correlation and relationship between these exposures and the outcome of both Type 1 and Type 2 diabetes. This data from the MRBase database was then analyzed to understand whether or not different exposure factors link to either type of diabetes.

Workflow Analysis

This study, as mentioned previously, involved the use of Mendelian Randomization and an MRBase database. First, a list of potential contributing exposures was loaded into the MRBase database to calculate their correlations with the outcome of Type 1 and 2 diabetes. This correlation was measured by calculating p-values. After the gathering of data, the numbers were then analyzed to come to certain conclusions about contributors towards both Type 1 and 2 diabetes. Mendelian Randomization, as mentioned earlier, prevented the incorporation of confounders in affecting the results of the data collected during the first stage.

Results

TABLE LEGEND	
Abbreviation	**Full Name**
WR	Wald Ratio
IVW	Inverse Variance Weighted
CE Size	Estimated Casual Effect Size
SE	Standard Error of Estmimated Casual Effect Size
P	P-Value

id.exposure	Outcome	Exposure	Method	SNPs (nsnp)	CE Size	SE	P
ukb-d-20117_2	T1D	Alcohol Drinker	WR	1	-3.17	4.62	0.49
ukb-d-20117_2	T2D	Alcohol Drinker	IVW	2	0.40	1.73	0.81
ukb-b-19953	T1D	BMI	IVW	36	-0.29	0.47	0.54
ukb-b-19953	T2D	BMI	IVW	346	1.07	0.04	0.00
ieu-a-9	T1D	Coronary HD	IVW	4	0.02	0.14	0.89
ieu-a-9	T2D	Coronary HD	IVW	9	0.04	0.06	0.53
met-c-859	T1D	Glucose	WR	1	-0.11	0.16	0.48
met-c-859	T2D	Glucose	IVW	2	0.19	0.42	0.66
ukb-b-10912	T1D	High Cholesterol	IVW	19	0.54	0.66	0.41
ukb-b-10912	T2D	High Cholesterol	IVW	49	1.19	1.04	0.25
ukb-b-3957	T1D	Insomnia	IVW	4	0.07	0.98	0.95
ukb-b-3957	T2D	Insomnia	IVW	32	0.31	0.23	0.18
ieu-a-1187	T1D	MDD	IVW	3	-0.06	0.47	0.89
ieu-a-1187	T2D	MDD	IVW	30	0.10	0.07	0.15
ukb-b-20261	T1D	Ever Smoked	IVW	6	3.88	6.33	0.54
ukb-b-20261	T2D	Ever Smoked	IVW	67	0.70	0.23	0.00
ieu-a-302	T1D	Triglycerides	IVW	16	-0.16	0.10	0.12
ieu-a-302	T2D	Triglycerides	IVW	51	0.03	0.09	0.70
ukb-b-14177	T1D	High BP	IVW	20	1.68	1.36	0.22
ukb-b-14177	T2D	High BP	IVW	170	1.18	0.19	0.00

For the analytical aspect of the research, a risk factors table was created through the use of R. After testing over twenty different exposures that potentially contribute to the development of Type 1 and Type 2 diabetes, defining exposures were found by analyzing p-values. The lower the p-value, the stronger the link between the exposure and outcome of Mendelian Randomization. Statistically speaking, a p-value between 0 and 1 is considered significant, and a p-value less than 0.05 is considered the most statistically significant. Causal effect size was also calculated, which measures the magnitude of the difference between the chosen exposures and outcomes. The p-value will illustrate that an outcome exists, but the causal effect size will reveal the size of the effect. The number of SNPs for each exposure is also mentioned in the table for each variant group because it illustrates which exposures and outcomes are more amplified than others. Hence, by calculating p-values and causal effect size, Type 1 and Type 2 diabetes were able to be tested for which exposures they are linked with.

The reason for testing both Type 1 and Type 2 diabetes separately for every risk factor is because they are different medical conditions. Hence, it was predicted in advance that they would likely not have the same contributing factors. By testing for every risk factor for each type of diabetes, separate p-values could be calculated to figure out the relationship between tested exposures and their impact on the outcome of diabetes in individuals. The usage of Mendelian Randomization to determine this relationship is also the best approach because in Mendelian Randomization analyses, variants can act as exposures in an environment independent of confounders. It is well known that confounding and biases can result in inaccurate results, which can be avoided through this method of analysis.

Type 2 Diabetes

Body Mass Index (BMI)

In the 2016 Corbin et al. study, body mass index (BMI) was tested as a modifiable risk factor for Type 2 diabetes. Specifically, they found 97 genetic variants of exposure within BMI that have a positive correlation with Type 2 diabetes [9]. A Mendelian Randomization of BMI and Type 2 diabetes with 361 genetic variants verified this causal relationship (Fig. 3), validating the Corbin et al. study and the Mendelian Randomization approach. My study in-

cludes more variants (361 variants), making my study better. As seen by the table, the p-value to demonstrate this relationship between the exposure and outcome is under 0.5, implying that the relationship is statistically significant. This makes sense because Type 2 diabetes is largely known to be caused by lifestyle choices. Thus, being overweight and having a higher BMI greatly increases one's risk of developing Type 2 diabetes because with a significant amount of weight, muscles become more resistant to insulin. However, my analysis of BMI with Type 1 Diabetes, shows no causal relationship (Fig. 4).

High Blood Pressure Diagnostic by Doctor

High blood pressure is often caused by a variety of factors, including obesity, certain types of diets high in fats, and having a sedentary lifestyle. The 2015 Emdin et al. study found that in a study of 4.1 million adults, the link between high blood pressure and Type 2 diabetes was undeniable [10]. They started off their study by recognizing that there are little to no studies that directly prove the relationship between high blood pressure and Type 2 diabetes using medical analysis, but through their examination of electronic health records, they found that a relationship without confounds does exist. The 4.1 million people they tested were adults who are free of diabetes and cardiovascular disease. A Mendelian Randomization of high blood pressure and Type 2 diabetes verified this causal relationship (Fig. 5), validating the Emdin et. al study and the Mendelian Randomization approach. As seen by the table generated by
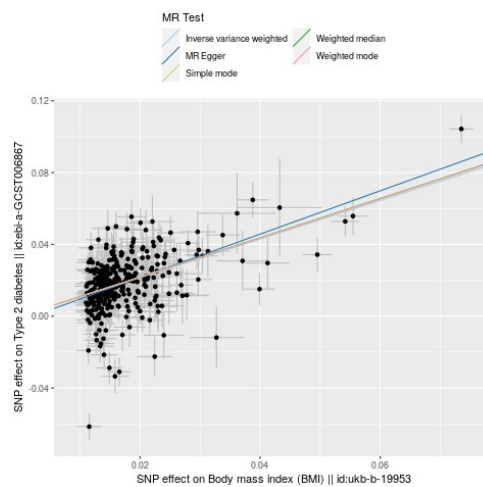

Fig. 3: *In this plot, Mendelian Randomization was used to analyze the effects of the exposure (BMI) on the outcome effect of Type 2 diabetes. Each data point represents one genetic variant and, based on this plot, one can see that with higher BMI, there is a positive correlation associated with getting Type 2 diabetes. This can logically be inferred considering upward position of the lines.*
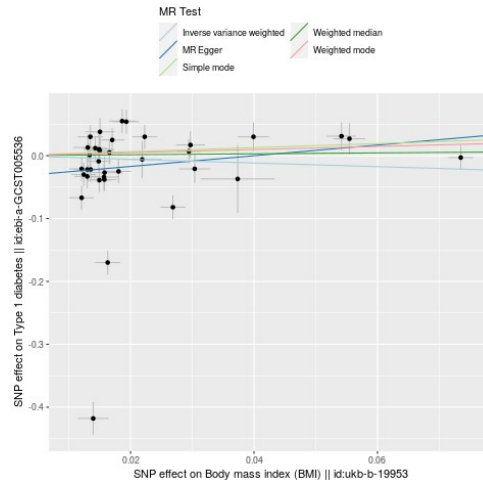
Fig. 4: *In this plot, Mendelian Randomization was used to analyze the effects of the exposure (BMI) on the outcome effect of Type 1 diabetes. Each data point represents one genetic variant and, based on this plot, one can see that with higher BMI, there is no correlation associated with getting Type 1 diabetes. This makes sense when considering all aspects of the plot.*

R, the p-value, while less statistically significant than BMI, proves a positive correlation between high blood pressure and Type 2 diabetes. While there may not be a plethora of medical knowledge behind this link, this can logically be assumed to be correct because high blood pressure can be caused by factors like living a sedentary lifestyle. This, in turn, increases BMI levels, which we know from the previous analysis is a statistically significant contributor towards Type 2 diabetes.
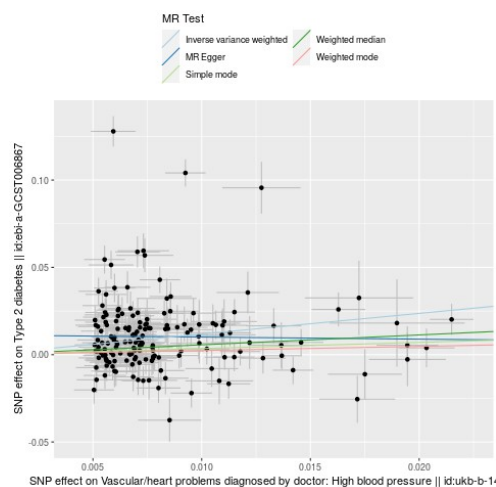

Fig. 5: *In this plot, Mendelian Randomization was used to analyze the effects of the exposure (high blood pressure) on the outcome effect of Type 2 diabetes. Each data point represents one genetic variant and, based on this plot, one can see that the results are inconclusive, as only the inverse variance weighted line displays a positive causal relationship between high blood cholesterol and Type 2 diabetes. The other Mendelian Randomizations, however, show no causal relationship between the exposure and the outcome.*

Smoking

In the 2009 Xie et al. study, smoking is acknowledged as a major contributor of Type 2 diabetes. The study specifically focuses on the impact of nicotine, which is an ingredient in many tobacco products, on the development of type 2 diabetes. The study found that smoking, nicotine in particular, "decreases insulin sensitivity and consequently results in the disorders of glucose and lipid metabolism" [11]. This implies that smoking harms metabolic control, which explains why smokers with Type 2 diabetes need higher insulin dosages than others. A Mendelian Randomization of smoking and Type 2 diabetes verified this relationship, validating the Xie et al. study and the Mendelian Randomization approach. While the p-value is less statistically significant than both BMI and high blood pressure, it is still statistically significant, as seen by the table generated by R. This logically makes sense because smokers are more likely to experience vascular conditions, which is caused by factors like high blood pressure. High blood pressure is a significant contributor of many vascular conditions, and as analyzed before, high blood pressure is a statistically significant contributor towards Type 2 diabetes.
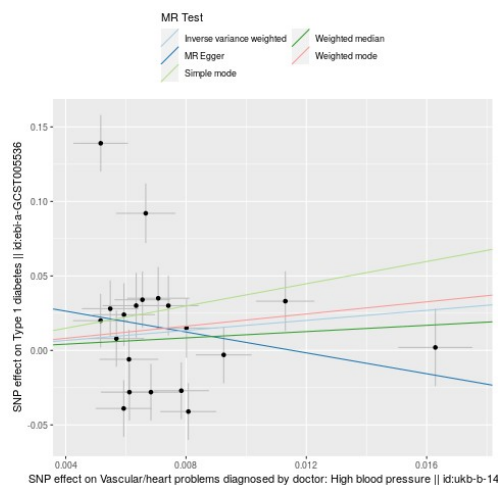

Fig. 6: In this plot, Mendelian Randomization was used to analyze the effects of the exposure (high blood pressure) on the outcome effect of Type 1 diabetes. Each data point represents one genetic variant and, based on this plot, there is a variation between the different Mendelian Randomizations. Hence, it can logically be concluded that there is no evidence for a causal relationship between high blood pressure and Type 1 diabetes.
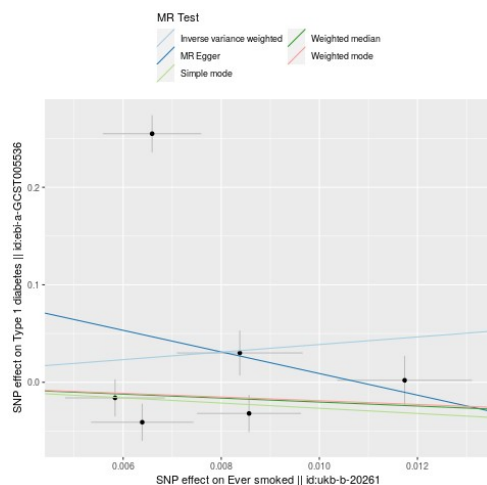

Fig. 7: In this plot, Mendelian Randomization was used to analyze the effects of the exposure (smoking) on the outcome effect of Type 1 diabetes. Each data point represents one genetic variant and, based on this plot, one can see

that there is a very small sample size, making the conclusion inconclusive. By the inverse variance weighted line, it can be seen that there is no p-value that is easily seen. This makes the graphed line not very significant. Hence, this line shows that there is a negative causal relationship between smoking and Type 1 diabetes. The MR Egger line, on the other hand, displays a positive causal relationship between the exposure and outcome, but, based on the table, there is no statistically significant p-value, making the relationship inconclusive. The other Mendelian Randomizations, however, display no causal relationship between the exposure and outcome.
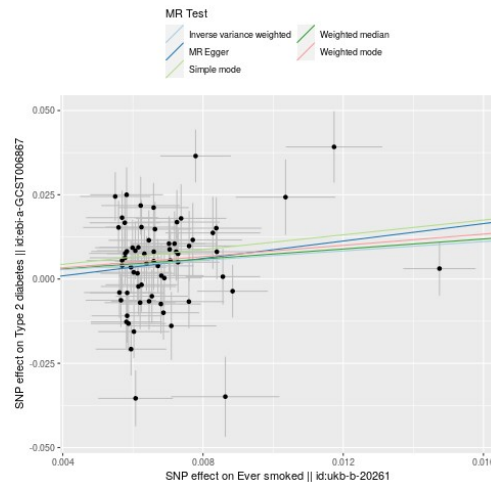
TYPE 1 DIABETES

Fig. 8: In this plot, Mendelian Randomization was used to analyze the effects of the exposure (smoking) on the outcome effect of Type 2 diabetes. Each data point represents one genetic variant and, based on this plot, it can be seen that there is a small sample size, making the p-value of the Mendelian Randomization prediction not statistically significant. However, the Mendelian Randomizations suggest a positive causal relationship between the exposure of smoking and the outcome of Type 2 diabetes, if the sample size was larger.

Unlike Type 2 diabetes, there are no statistically significant relationships between exposures and the onset of Type 1 diabetes. This could potentially be because Type 1 diabetes is statistically more likely to occur in younger individuals. This type of diabetes specifically occurs when there is a genetic mutation that causes pancreatic cells in the body to be destroyed. Type 1 diabetes in not known to be caused by lifestyle choices, as of studies conducted up until now. Using Mendelian Randomization, this idea can be verified, as there is no statistically significant link between the exposures (BMI, high blood pressure, and smoking) and the outcome of Type 1 diabetes. However, certain factors and choices made in this study may have affected this conclusion. The dataset for Type 1 diabetes only included 6,683 people, while the date set for Type 2 diabetes included 61,714 people – around ten times larger. Hence, Type 2 diabetes, being one with more samples in this study, would have more statistical power to identify causal associations with Mendelian Randomization.

Discussion

The analytical study presented is imperative in analyzing a variety of Mendelian Randomization factors to prove or disprove a link between itself and either type of diabetes. Diabetes is one of the leading cause of death for individuals internationally [1], making research on its exposures even more critical. As seen from the results section, there is a common defining exposure that causes Type 2 diabetes, but there is no such defining exposure that is linked to

Type 1 diabetes. Specifically, the three defining contributors of Type 2 diabetes found through this study are body mass index (BMI), high blood pressure, and smoking. For Type 1 diabetes, there were no defining contributing exposures found in this study because Type 1 diabetes is not statistically known to be caused by external factors.

A lot of the contributing exposures to Type 2 diabetes could potentially be due to the effects of housing density. It is important to note that housing density does not have a strong genetic component. However, with more house members, there are more responsibilities, leading to higher stress levels. Therefore, it is imperative to analyze factors like stress and unhealthy behaviors such as insomnia and the type of diet followed. Higher levels of stress are known to be associated with high blood pressure [12], which, as analyzed previously, has a statistically significant link to Type 2 diabetes. The type of diet a person follows can be influenced by peers, and having unhealthy diets can lead to obesity, especially in poorer countries with high poverty rates, where family sizes are generally known to be much higher than in more developed countries [13]. Obesity, as mentioned in the results section, causes high BMI levels, which is the most statistically significant contributor of Type 2 diabetes.

Not only this, but factors controlled by governments can also contribute to an individual being more likely to develop Type 2 diabetes. As seen in the 2012 Hsu et al. study, inequalities of health care coverage in many international countries can also contribute to a more likely outcome of developing Type 2 diabetes. Specifically, the study analyzed that with a lack of health care coverage, residents living in these countries are less likely to pay for medications, testing, and medical assistance due to high charges [14]. This, unfortunately, results in these residents being more likely to remain untreated for major health concerns, especially issues like vascular conditions, which play a significant role in the development of Type 2 diabetes.

When analyzing both housing effects and the effects of health care coverage in countries, there was no link found between these exposures and the outcome of Type 1 diabetes. However, it can logically be assumed that due to a lack of health care coverage or by living in a household with many people, an individual's medical concerns will be less efficiently catered to, as compared to treatments for individuals living in developed countries. Due to this, there will be much worse and severe cases of both types of diabetes, especially Type 1 diabetes, due to a lack of the necessary amount of insulin dosages. This, unfortunately, links to higher mortality rates.

References

"The top 10 causes of death," *World Health Organization*, 2018.

M. B. Nasr, S. Tezza, F. D'Addio, C.   Mameli, V. Usuelli,
Maestroni, D. Corradi, S. Belletti, L. Albarello, G. Becchi, G. P. Fadini, C. Schuetz, J. Markmann, C. Wasserfall, L. Zon, G. V. Zuccotti, and P. Fiorina, "PD-l1 genetic overexpression or pharmacological restoration in hematopoietic stem and progenitor cells reverses autoimmune diabetes," *Science Translational Medicine*, vol. 9, p. eaam7543, Nov. 2017.

"About diabetes," *World Health Organization*, 2014.

M. Monaghan, V. Helgeson, and D. Wiebe, "Type 1 diabetes in young adulthood," *Current Diabetes Reviews*, vol. 11, pp. 239–250, July 2015.

S. Chatterjee, K. Khunti, and M. J. Davies, "Type 2 diabetes," *The Lancet*, vol. 389, pp. 2239–2251, June 2017.

A. Bianco, F. Pomara, E. Thomas, A. Paoli, G. Battaglia, M. Petrucci, P. Proia, M. Bellafiore, and A. Palma, "Type 2 diabetes family histories, body composition and fasting glucose levels: a cross-section analysis in healthy sedentary male and female," *Iran. J. Public Health*, vol. 42, no. 7, pp. 681–690, 2013.

D. Dayal, "Understanding diabetes from a new perspective: The role of free radicals," *Free Radicals in Biology and Medicine*, vol. 77, 2005.

M. V. Holmes, M. Ala-Korpela, and G. D. Smith, "Mendelian randomization in cardiometabolic disease: challenges in evaluating causality," *Nature Reviews Cardiology*, vol. 14, pp. 577–590, June 2017.

L. J. Corbin, R. C. Richmond, K. H. Wade, S. Burgess, J. Bowden, G. D. Smith, and N. J. Timpson, "BMI as a modifiable risk factor for type 2 diabetes: Refining and understanding causal estimates using mendelian randomization," *Diabetes*, vol. 65, pp. 3002–3007, July 2016.

C. A. Emdin, S. G. Anderson, M. Woodward, and K. Rahimi, "Usual blood pressure and risk of new-onset diabetes," *Journal of the American College of Cardiology*, vol. 66, pp. 1552–1562, Oct. 2015.

X. tao Xie, Q. Liu, J. Wu, and M. Wakui, "Impact of cigarette smoking in type 2 diabetes development," *Acta Pharmacologica Sinica*, vol. 30, pp. 784–787, May 2009.

P. Suter, R. Maire, D. Holtz, and W. Vetter, "Relationship between self-perceived stress and blood pressure," *Journal of Human Hypertension*, vol. 11, pp. 171–176, Mar. 1997.

H. A. Lee, W. K. Lee, K.-A. Kong, N. Chang, E.-H. Ha, Y. S. Hong, and H. Park, "The effect of eating behavior on being overweight or obese during preadolescence," *Journal of Preventive Medicine and Public Health*, vol. 44, pp. 226–233, Sept. 2011.

C.-C. Hsu, C.-H. Lee, M. L. Wahlqvist, H.-L. Huang, H.-Y. Chang, L. Chen, S.-F. Shih, S.-J. Shin, W.-C. Tsai, T. Chen, C.-T. Huang, and J.-S. Cheng, "Poverty increases type 2 diabetes incidence and inequality of care despite universal health coverage," *Diabetes Care*, vol. 35, pp. 2286–2292, Aug. 2012.