

# Modulation of Diabetes Risk by Lifestyle Behaviors in the Presence of Elevated Body Mass Index

data-to-paper

April 17, 2024

## Abstract

Diabetes constitutes a growing public health concern, exacerbated by lifestyle behaviors in conjunction with increased body mass index (BMI). Addressing the critical need to understand how modifiable lifestyle factors can influence diabetes risk, our research investigates the complex interplay of physical activity, smoking, and dietary habits on the BMI-diabetes association. Leveraging a robust dataset from the CDC's Behavioral Risk Factor Surveillance System, this study encompasses 253,680 American adults. We employed statistical models to parse the moderation effects of lifestyle behaviors on the risk of diabetes as influenced by BMI. Our key findings articulate that regular physical activity mitigates the risk of diabetes even in the presence of high BMI, whereas smoking amplifies this risk. Dietary analyses further reveal that high vegetable consumption is associated with a reduced effect of BMI on diabetes, a benefit not observed with fruit intake. Importantly, the study maps the influence of socioeconomic parameters on these associations, underpinning their epidemiological relevance. Whilst the cross-sectional nature of our data restricts causal inference, and self-reported measures necessitate circumspect interpretation, the evidence affirms the potency of lifestyle modifications in diabetes intervention strategies. These insights bridge a gap in understanding and bear significant implications for public health policy and individual behavior change frameworks.

## Introduction

The specter of diabetes looms large over public health, with its prevalence doubling over the past two decades and accounting for substantial morbidity and mortality worldwide [1, 2]. The disease's pervasive nature is exacerbated by modifiable lifestyle factors, including diet, physical activity,

and smoking habits, especially in the context of the body mass index (BMI) [3, 4, 5]. While the deleterious impact of increased BMI on diabetes risk is well-documented, the "complex interplay" involving the mitigation or exacerbation of this risk by individual lifestyle choices remains inadequately explicated [6, 7].

Prior research has predominantly focused on the independent effects of lifestyle behaviors on diabetes risk, leaving a knowledge gap regarding their collective modulation effects on the obesity-diabetes continuum, especially given the heterogeneous nature of diabetes manifestations across various populations [8, 9, 10]. Studies highlighting the beneficial role of physical activity and dietary patterns have paved the way for lifestyle interventions, but the nuance of interaction effects between such behaviors and BMI, particularly concerning fruit and vegetable consumption, has been less explored and warrants further study [11, 12, 13]. Similarly, while the adverse effects of smoking on diabetes risk are established, its specific interaction with BMI is still ill-defined, leaving questions about quantifying such a moderation effect [14, 15].

This research endeavors to fill these gaps, utilizing the expansive and detailed data from the BRFSS, which provides insights into the health-related behaviors of American adults and their impact on chronic diseases, including diabetes [16, 17, 18]. By integrating these rich datasets, our study investigates how lifestyle behaviors modulate the effects of BMI on diabetes risk, providing empirical evidence of whether physical activity, dietary habits, and smoking serve to either mitigate or exacerbate this association.

Methodologically, our analysis employs ordinary least squares (OLS) regression—a technique well-suited to examining the interdependent effects of multiple variables—thereby allowing us to capture the nuanced moderation played by lifestyle behaviors on the BMI-diabetes nexus [19, 20, 21]. Our results yield statistically significant insights, elucidating how physical activity and vegetable consumption can weaken, while smoking can enhance, the diabetes-inducing potential of high BMI. These findings contribute to forming an interconnected picture that informs individual behavioral and public health policies aimed at diabetes prevention and control [22, 23].

## Results

First, to characterize the sample in terms of diabetes and associated lifestyle behaviors, we analyzed descriptive statistics for key variables. As shown in Table 1, the average Body Mass Index (BMI) was 28.4 with a standard

deviation of 6.61. The prevalence of diabetes was observed to be 13.9%, with 75.7% of the participants engaging in physical activity ('PhysActivity') in the past 30 days. The proportion of individuals classified as smokers ('Smoker') was 44.3%. Daily fruit ('Fruits') and vegetable ('Veggies') consumption was reported at 63.4% and 81.1%, respectively. These descriptive findings draw from a dataset encompassing a total of 253680 observations.

Table 1: Descriptive statistics of key variables

	mean	std
<b>Diabetes</b>	0.139	0.346
<b>BMI</b>	28.4	6.61
<b>Physical Activity</b>	0.757	0.429
<b>Smoker</b>	0.443	0.497
<b>Fruits</b>	0.634	0.482
<b>Veggies</b>	0.811	0.391

NOTE: The number of observations in all variables is 253680.0

**Diabetes:** Diabetes occurrence. 1 if yes, 0 otherwise

**Physical Activity:** Phys. Activity in past 30 days, 1: Yes, 0: No

Then, to examine whether physical activity influences the association between BMI and diabetes, regression analysis was conducted incorporating an interaction term between BMI and physical activity ('PhysActivity'). The significant interaction term ( $-0.00221$ ;  $p < 10^{-6}$ ) from Table 2 suggests that physical activity moderates the relationship between BMI and diabetes risk. Specifically, the increase in the odds of having diabetes per unit increase in BMI is attenuated for physically active individuals by  $-0.002208$ , indicating a reduction in diabetes risk.

To further assess lifestyle factors, we investigated the impact of smoking ('Smoker') on the BMI-diabetes link. As reported in Table 3, the interaction of BMI and smoking status was significant ( $0.00273$ ;  $p < 10^{-6}$ ), indicating that smoking modifies the impact of BMI on diabetes prevalence. The risk of diabetes associated with each incremental increase in BMI was heightened for smokers, demonstrated by an increased risk multiplier of  $1.003$  compared to non-smokers, suggesting that smoking amplifies the diabetes risk associated with a higher BMI.

Finally, the effect of diet was assessed by analyzing the interaction of BMI with daily fruit and vegetable consumption ('Fruits' and 'Veggies'). The results from Table 4 show a statistically significant interaction term for vegetables and BMI ( $-0.000577$ ;  $p = 0.0223$ ), illustrating a reduction in

Table 2: Analysis of relationship between BMI and Diabetes moderated by Physical Activity

	Coef.	Std.Err.	t-val	p-val	[0.025	0.975]
<b>Intercept</b>	-0.175	0.00678	-25.8	$<10^{-6}$	-0.188	-0.162
<b>BMI</b>	0.012	0.000175	68.5	$<10^{-6}$	0.0116	0.0123
<b>Physical Activity</b>	0.0266	0.00647	4.1	$4.06 \cdot 10^{-5}$	0.0139	0.0392
<b>BMI * Phys. Act.</b>	-0.00221	0.000213	-10.4	$<10^{-6}$	-0.00263	-0.00179
<b>Age</b>	0.0187	0.000216	86.5	$<10^{-6}$	0.0183	0.0191
<b>Gender</b>	0.0305	0.00133	23	$<10^{-6}$	0.0279	0.0331
<b>Education</b>	-0.0112	0.000749	-14.9	$<10^{-6}$	-0.0127	-0.00971
<b>Income</b>	-0.0176	0.00036	-48.8	$<10^{-6}$	-0.0183	-0.0169

**Age:** 13-level age category in intervals of 5 years (e.g., 1 = 18-24, 2 = 25-29)

**Gender:** 1 if male, 0 if female

**Education:** Education Level. 1-6 with 1 being "Never attended school" and 6 being "College Graduate"

**Income:** Income Scale. 1-8 with 1 being " $\leq \$10K$ " and 8 being " $> \$75K$ "

**t-val:** t-statistic of the regression estimate

**p-val:** Probability that the null hypothesis (of no relationship) produces results as extreme as the estimate

**BMI \* Phys. Act.:** Interaction between BMI and Physical Activity

**Physical Activity:** Phys. Activity in past 30 days, 1: Yes, 0: No

the odds ratio of diabetes risk per unit increase in BMI for individuals who consume vegetables daily by -0.0005768. In contrast, fruit consumption did not have a statistically significant moderating effect on the BMI-diabetes relationship (0.000144;  $p = 0.492$ ).

In summary, these results reveal that the interaction between physical activity and BMI is significantly associated with a reduced risk of diabetes, while smoking is associated with an increased risk. Furthermore, vegetable consumption is associated with a slight but statistically significant reduction in diabetes risk in the context of higher BMI levels, but no such moderating influence is observed for fruit consumption.

## Discussion

This study contributes to a more profound understanding of diabetes—a prominent global health issue—in the context of modifiable lifestyle behaviors and high body mass index (BMI) [1, 4]. Our investigation relied on an extensive dataset provided by the Centers for Disease Control and Pre-

Table 3: Analysis of relationship between BMI and Diabetes moderated by Smoking Status

	Coef.	Std.Err.	t-val	p-val	[0.025	0.975]
<b>Intercept</b>	-0.129	0.00587	-22	$<10^{-6}$	-0.14	-0.117
<b>BMI</b>	0.00962	0.000132	72.7	$<10^{-6}$	0.00936	0.00988
<b>Smoker</b>	-0.0675	0.00583	-11.6	$<10^{-6}$	-0.0789	-0.0561
<b>BMI * Smoker</b>	0.00273	0.000199	13.7	$<10^{-6}$	0.00234	0.00312
<b>Age</b>	0.019	0.000217	87.4	$<10^{-6}$	0.0186	0.0194
<b>Gender</b>	0.0282	0.00134	21.1	$<10^{-6}$	0.0256	0.0308
<b>Education</b>	-0.0125	0.000749	-16.7	$<10^{-6}$	-0.014	-0.0111
<b>Income</b>	-0.0184	0.000359	-51.2	$<10^{-6}$	-0.0191	-0.0177

**Age:** 13-level age category in intervals of 5 years (e.g., 1 = 18-24, 2 = 25-29)

**Gender:** 1 if male, 0 if female

**Education:** Education Level. 1-6 with 1 being "Never attended school" and 6 being "College Graduate"

**Income:** Income Scale. 1-8 with 1 being " $\leq \$10K$ " and 8 being " $> \$75K$ "

**t-val:** t-statistic of the regression estimate

**p-val:** Probability that the null hypothesis (of no relationship) produces results as extreme as the estimate

**Smoker:** 1 if smoker, 0 otherwise

**BMI \* Smoker:** Interaction between BMI and Smoking

vention's Behavioral Risk Factor Surveillance System (BRFSS), focusing on the effects of physical activity, smoking, and dietary habits on the frequently observed association between high BMI and diabetes [24, 3].

Applying ordinary least squares regression models and incorporating interaction terms, we evaluated the moderation effect of different lifestyle behaviors on the diabetes risk associated with BMI [19, 21, 23]. Our primary findings underscore that regular physical activity can mitigate the heightened risk of diabetes connected with increased BMI, aligning with research emphasizing the importance of physical activity in managing diabetes and extending it by highlighting its potential in diabetes prevention [11, 25]. Furthermore, smoking was discovered to exacerbate the BMI-diabetes relationship, reinforcing the need for anti-smoking efforts in diabetes prevention strategies [14, 15].

Significantly, our data revealed that high vegetable consumption desensitizes the effect of BMI on diabetes risk, a finding that marries well with the extant literature advocating for diet-focused interventions against diabetes [13, 26]. However, we found no similar moderating influence for fruit

Table 4: Analysis of relationship between BMI and Diabetes moderated by Consumption of Fruits and Vegetables

	Coef.	Std.Err.	t-val	p-val	[0.025	0.975]
<b>Intercept</b>	-0.155	0.00797	-19.5	$<10^{-6}$	-0.171	-0.14
<b>BMI</b>	0.0111	0.000228	48.7	$<10^{-6}$	0.0107	0.0116
<b>Fruits</b>	-0.0143	0.00618	-2.32	0.0206	-0.0264	-0.00219
<b>BMI * Fruits</b>	0.000144	0.00021	0.687	0.492	-0.000267	0.000555
<b>Veggies</b>	0.004	0.00754	0.531	0.595	-0.0108	0.0188
<b>BMI * Veggies</b>	-0.000577	0.000252	-2.28	0.0223	-0.00107	$-8.19 \cdot 10^{-5}$
<b>Age</b>	0.0192	0.000217	88.7	$<10^{-6}$	0.0188	0.0196
<b>Gender</b>	0.0276	0.00134	20.6	$<10^{-6}$	0.025	0.0302
<b>Education</b>	-0.0121	0.000749	-16.1	$<10^{-6}$	-0.0136	-0.0106
<b>Income</b>	-0.0181	0.00036	-50.2	$<10^{-6}$	-0.0188	-0.0174

**Age:** 13-level age category in intervals of 5 years (e.g., 1 = 18-24, 2 = 25-29)

**Gender:** 1 if male, 0 if female

**Education:** Education Level. 1-6 with 1 being "Never attended school" and 6 being "College Graduate"

**Income:** Income Scale. 1-8 with 1 being " $\leq \$10K$ " and 8 being " $> \$75K$ "

**t-val:** t-statistic of the regression estimate

**p-val:** Probability that the null hypothesis (of no relationship) produces results as extreme as the estimate

**Fruits:** One fruit/day, 1: Yes, 0: No

**Veggies:** One veggie/day, 1: Yes, 0: No

**BMI \* Fruits:** Interaction between BMI and Fruit consumption

**BMI \* Veggies:** Interaction between BMI and Vegetable consumption

consumption. This distinction prompts a reconsideration of the role of different types of dietary habits in managing diabetes and encourages further exploration into the reasons for the apparent discrepancy in the effects of vegetable and fruit consumption.

Despite our findings' significance, we acknowledge some limitations. Firstly, our use of a cross-sectional study design impedes our ability to infer causal relationships. Furthermore, the self-reported nature of the variables, inherent to the BRFSS data set, presents a potential recall bias that might afflict some of the respondents, leading to potential inaccuracies in the reported data. Future research employing longitudinal study designs and objective measurements can help affirm these findings and better establish causal inferences.

Our study fundamentally underscores that lifestyle modifications hold considerable promise for mitigating diabetes risk in the context of high

BMI. While we recognize the differing effects of individual behaviors, it is essential to consider them collectively when designing intervention strategies for diabetes prevention [10]. Moreover, our results particularly urge reconsideration of current dietary guidelines, hinting that promoting vegetable consumption might offer more substantial benefits in diabetes management than fruit intake.

Although our analysis spans a broad demographic base, our findings might bear specific importance for populations at higher risk of diabetes, including older adults, certain ethnic groups, and economically disadvantaged individuals. Future research can unravel these associations more intimately across various demographic groups to guide more targeted interventions.

In conclusion, this research unravels the complex interplay of lifestyle behaviors and BMI in diabetes risk, and the findings bear crucial implications for both individual lifestyle adjustments and broader public health efforts. Our study elegantly illustrates how behavioral modifications are not only potent prevention strategies against diabetes but also critical for managing the increased diabetes risk associated with high BMI. Future research should strive to corroboratively substantiate these results, delve deeper into the mechanisms underlying these relationships, and proactively shape policy and practice aiming to curb the overwhelming burden of diabetes.

## Methods

### Data Source

The dataset employed in this study originates from the Centers for Disease Control and Prevention’s (CDC) Behavioral Risk Factor Surveillance System (BRFSS) for the year 2015, a comprehensive telephone-based survey that collects annual data on health-related risk behaviors, chronic health conditions, and the utilization of preventative services from American adults. The dataset included 253,680 participants, with each individual being represented as a data point across 22 features. These encompassed demographic information, lifestyle choices, and health indicators pertinent to the examination of diabetes risk. Missing values had previously been expunged, yielding a dataset ready for robust statistical analyses without further need for imputation.

## Data Preprocessing

As part of the analysis, initial preparations to the dataset were not required since the dataset was already in a clean and usable format with no missing values. Consequently, our study proceeded directly with the application of statistical techniques to the extant dataset as provided.

## Data Analysis

In assessing the relationships embedded within our data, we pursued multiple regression analyses with the application of ordinary least squares (OLS) methodology. The primary focus of these analyses was to discern the moderation effect of various lifestyle behaviors on the relationship between Body Mass Index (BMI) and diabetes incidence. To this end, we constructed several regression models each evaluating BMI in conjunction with a distinct lifestyle variable—specifically, physical activity, smoking status, and fruit and vegetable consumption—and their interaction terms.

The first model integrated the interaction between BMI and physical activity alongside control variables. Subsequent models similarly considered the interactions between BMI and smoking, and BMI and both dietary behaviors, respectively, each inclusive of the same control variables for age, sex, educational attainment, and income level, to ensure consistency and comparability across models.

The results from each fitted model were detailed with comprehensive statistical outputs, including coefficients, standard errors, and significance levels. These outputs furnished us with the insights necessary to evaluate our original hypothesis and to elucidate the potential moderating roles of the examined lifestyle behaviors on diabetes risk in relation to BMI.

## Code Availability

Custom code used to perform the data preprocessing and analysis, as well as the raw code outputs, are provided in Supplementary Methods.

## References

- [1] M. Chiu, P. Austin, D. Manuel, B. Shah, and J. Tu. Deriving ethnic-specific bmi cutoff points for assessing diabetes risk. *Diabetes Care*, 34:1741 – 1748, 2011.



- [2] Gitanjali M Singh, G. Danaei, F. Farzadfar, Gretchen A. Stevens, M. Woodward, D. Wormser, S. Kaptoge, G. Whitlock, Q. Qiao, S. Lewington, E. Di Angelantonio, S. Vander Hoorn, C. Lawes, Mohammed K. Ali, D. Mozaffarian, and M. Ezzati. The age-specific quantitative effects of metabolic risk factors on cardiovascular diseases and diabetes: A pooled analysis. *PLoS ONE*, 8, 2013.
- [3] J. Chan, E. Rimm, G. Colditz, M. Stampfer, and W. Willett. Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. *Diabetes Care*, 17:961 – 969, 1994.
- [4] Ying Chen, Xiao-Ping Zhang, Jie Yuan, Bo zhi Cai, Xiao-Li Wang, Xiao li Wu, Yueqi Zhang, Xiao-Yi Zhang, T. Yin, Xiaohui Zhu, Y. Gu, S. Cui, Zhidong Lu, and Xiao ying Li. Association of body mass index and age with incident diabetes in chinese adults: a population-based cohort study. *BMJ Open*, 8, 2018.
- [5] Natallia Gray, G. Picone, F. Sloan, and A. Yashkin. Relation between bmi and diabetes mellitus and its complications among us older adults. *Southern Medical Journal*, 108:2936, 2015.
- [6] X. Pi-Sunyer, G. Blackburn, F. Brancati, G. Bray, R. Bright, J. Clark, J. M. Curtis, M. Espeland, J. Foreyt, Kathryn Graves, S. Haffner, B. Harrison, James O Hill, E. Horton, J. Jakicic, R. Jeffery, K. Johnson, S. Kahn, D. Kelley, A. Kitabchi, W. Knowler, C. Lewis, B. J. Maschak-Carey, B. Montgomery, D. Nathan, Jennifer Patricio, A. Peters, J. Redmon, R. Reeves, D. Ryan, M. Safford, B. van Dorsten, T. Wadden, L. Wagenknecht, Jacqueline Wesche-Thobaben, R. Wing, and S. Yanovski. Reduction in weight and cardiovascular disease risk factors in individuals with type 2 diabetes. *Diabetes Care*, 30:1374 – 1383, 2007.
- [7] R. Wing, W. Lang, T. Wadden, M. Safford, W. Knowler, A. Bertoni, James O Hill, F. Brancati, A. Peters, and L. Wagenknecht. Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes. *Diabetes Care*, 34:1481 – 1486, 2011.
- [8] J. Lv, Canqing Yu, Yu Guo, Z. Bian, Ling Yang, Yiping Chen, Ximin Hu, W. Hou, Junshi Chen, Zhengming Chen, L. Qi, and Liming Li. Adherence to a healthy lifestyle and the risk of type 2 diabetes in chinese adults. *International Journal of Epidemiology*, 46:1410 – 1420, 2017.

- [9] T. Schnurr, Hermina Jakupovi, Germn D. Carrasquilla, L. ngquist, N. Grarup, T. Srensen, A. Tjnneland, K. Overvad, O. Pedersen, T. Hansen, and T. Kilpelinen. Obesity, unfavourable lifestyle and genetic risk of type 2 diabetes: a case-cohort study. *Diabetologia*, 63:1324–1332, 2020.
- [10] A. Bernab-Ortiz, R. Carrillo-Larco, R. Gilman, W. Checkley, L. Smeeth, J. Miranda, Cronicas Cohort, and Study Group. Contribution of modifiable risk factors for hypertension and type-2 diabetes in peruvian resource-limited settings. *Journal of Epidemiology and Community Health*, 70:49 – 55, 2015.
- [11] B. Bohn, A. Herbst, M. Pfeifer, D. Krakow, S. Zimny, F. Kopp, A. Melmer, J. Steinacker, and R. Holl. Impact of physical activity on glycemic control and prevalence of cardiovascular risk factors in adults with type 1 diabetes: A cross-sectional multicenter study of 18,028 patients. *Diabetes Care*, 38:1536 – 1543, 2015.
- [12] Xu Han, Yue Wei, Hua Hu, Jing Wang, Z. Li, Fei Wang, Tengfei Long, Jing Yuan, P. Yao, Sheng Wei, Youjie Wang, Xiaomin Zhang, Huan Guo, Handong Yang, Tangchun Wu, and M. He. Genetic risk, a healthy lifestyle, and type 2 diabetes: the dongfeng-tongji cohort study. *The Journal of clinical endocrinology and metabolism*, 2020.
- [13] M. Delgado-Velandia, Vannina Gonzlez-Marrachelli, A. Domingo-Relloso, M. Glvez-Fernndez, M. Grau-Prez, P. Olmedo, I. Galn, F. RodrguezArtalejo, N. Amig, L. Briongos-Figuero, J. Redn, J. Martn-Escudero, Daniel Monlen-Salvado, M. Tellez-Plaza, and M. Sotos-Prieto. Healthy lifestyle, metabolomics and incident type 2 diabetes in a population-based cohort from spain. *The International Journal of Behavioral Nutrition and Physical Activity*, 19, 2022.
- [14] R. Ng, R. Sutradhar, Z. Yao, W. Wodchis, and L. Rosella. Smoking, drinking, diet and physical activitymodifiable lifestyle risk factors and their associations with age to first chronic disease. *International Journal of Epidemiology*, 49:113 – 130, 2019.
- [15] Liang Shi, X. Shu, Hong-Lan Li, H. Cai, Qiaolan Liu, W. Zheng, Y. Xiang, and R. Villegas. Physical activity, smoking, and alcohol consumption in association with incidence of type 2 diabetes among middle-aged and elderly chinese men. *PLoS ONE*, 8, 2013.

- [16] Kevin A. Matthews, J. Croft, Yong Liu, Hua Lu, D. Kanny, A. Wheaton, T. Cunningham, L. Khan, R. Caraballo, J. Holt, P. Eke, and W. Giles. Health-related behaviors by urban-rural county classification united states, 2013. *MMWR Surveillance Summaries*, 66:1 – 8, 2017.
- [17] Ronaldo Iachan, Carol Pierannunzi, Kristie Healey, K. Greenlund, and M. Town. National weighting of data from the behavioral risk factor surveillance system (brfss). *BMC Medical Research Methodology*, 16, 2016.
- [18] Elizabeth L. Tung, Arshiya A Baig, E. Huang, N. Laiteerapong, and Kao-Ping Chua. Racial and ethnic disparities in diabetes screening between asian americans and other adults: Bfss 20122014. *Journal of General Internal Medicine*, 32:423–429, 2017.
- [19] A. Sambola, J. Osende, J. Hathcock, Michael Degen, Y. Nemerson, V. Fuster, J. Crandall, and J. Badimn. Role of risk factors in the modulation of tissue factor activity and blood thrombogenicity. *Circulation: Journal of the American Heart Association*, 107:973–977, 2003.
- [20] S. A. Black, K. Markides, and L. Ray. Depression predicts increased incidence of adverse health outcomes in older mexican americans with type 2 diabetes. *Diabetes care*, 26 10:2822–8, 2003.
- [21] Y. Rochlani, N. Pothineni, S. Kovelamudi, and J. Mehta. Metabolic syndrome: pathophysiology, management, and modulation by natural compounds. *Therapeutic Advances in Cardiovascular Disease*, 11:215 – 225, 2017.
- [22] S. Reutrakul, Megan M. Hood, S. Crowley, M. K. Morgan, M. Teodori, K. Knutson, and E. van Cauter. Chronotype is independently associated with glycemic control in type 2 diabetes. *Diabetes Care*, 36:2523 – 2529, 2013.
- [23] M. Stumvoll, A. Mitrakou, W. Pimenta, T. Jenssen, H. Yki-Jrvinen, T. W. Haeften, W. Renn, and J. Gerich. Use of the oral glucose tolerance test to assess insulin release and insulin sensitivity. *Diabetes care*, 23 3:295–301, 2000.
- [24] J. F. Pearson, Chethan Bachiredy, Sangameswaran Shyamprasad, A. Goldfine, and J. Brownstein. Association between fine particulate

matter and diabetes prevalence in the united states running title: Association of particulate matter and diabetes. 2010.

- [25] J. Reis, C. Loria, P. Sorlie, Yikyung Park, A. Hollenbeck, and A. Schatzkin. Lifestyle factors and risk for new-onset diabetes. *Annals of Internal Medicine*, 155:292 – 299, 2011.
- [26] Yanping Li, Dong D. Wang, S. Ley, Malik Vasanti, A. Howard, Yu na He, and F. Hu. Time trends of dietary and lifestyle factors and their potential impact on diabetes burden in china. *Diabetes Care*, 40:1685 – 1694, 2017.

## A Data Description

Here is the data description, as provided by the user:

The dataset includes diabetes related factors extracted from the CDC's Behavioral Risk Factor Surveillance System (BRFSS), year 2015.

The original BRFSS, from which this dataset is derived, is a health-related telephone survey that is collected annually by the CDC.

Each year, the survey collects responses from over 400,000 Americans on health-related risk behaviors, chronic health conditions, and the use of preventative services. These features are either questions directly asked of participants, or calculated variables based on individual participant responses.

1 data file:

"diabetes\\_binary\\_health\\_indicators\\_BRFSS2015.csv"

The csv file is a clean dataset of 253,680 responses (rows) and 22 features (columns).

All rows with missing values were removed from the original dataset; the current file contains no missing values.

The columns in the dataset are:

\#1 'Diabetes\\_binary': (int, bool) Diabetes (0=no, 1=yes)  
\#2 'HighBP': (int, bool) High Blood Pressure (0=no, 1=yes)  
\#3 'HighChol': (int, bool) High Cholesterol (0=no, 1=yes)  
\#4 'CholCheck': (int, bool) Cholesterol check in 5 years (0=no, 1=yes)  
\#5 'BMI': (int, numerical) Body Mass Index  
\#6 'Smoker': (int, bool) (0=no, 1=yes)  
\#7 'Stroke': (int, bool) Stroke (0=no, 1=yes)  
\#8 'HeartDiseaseorAttack': (int, bool) coronary heart disease (CHD) or myocardial infarction (MI), (0=no, 1=yes)  
\#9 'PhysActivity': (int, bool) Physical Activity in past 30 days (0=no, 1=yes)  
\#10 'Fruits': (int, bool) Consume one fruit or more each day (0=no, 1=yes)  
\#11 'Veggies': (int, bool) Consume one Vegetable or more each day (0=no, 1=yes)  
\#12 'HvyAlcoholConsump': (int, bool) Heavy drinkers (0=no, 1=yes)

```

\#13 'AnyHealthcare' (int, bool) Have any kind of health care
      coverage (0=no, 1=yes)
\#14 'NoDocbcCost' (int, bool) Was there a time in the past 12
      months when you needed to see a doctor but could not
      because of cost? (0=no, 1=yes)
\#15 'GenHlth' (int, ordinal) self-reported health (1=excellent
      , 2=very good, 3=good, 4=fair, 5=poor)
\#16 'MentHlth' (int, ordinal) How many days during the past 30
      days was your mental health not good? (1 - 30 days)
\#17 'PhysHlth' (int, ordinal) Hor how many days during the
      past 30 days was your physical health not good? (1 - 30
      days)
\#18 'DiffWalk' (int, bool) Do you have serious difficulty
      walking or climbing stairs? (0=no, 1=yes)
\#19 'Sex' (int, categorical) Sex (0=female, 1=male)
\#20 'Age' (int, ordinal) Age, 13-level age category in
      intervals of 5 years (1= 18 - 24, 2= 25 - 29, ..., 12= 75 -
      79, 13 = 80 or older)
\#21 'Education' (int, ordinal) Education level on a scale of 1
      - 6 (1=Never attended school, 2=Elementary, 3=Some high
      school, 4=High school, 5=Some college, 6=College)
\#22 'Income' (int, ordinal) Income scale on a scale of 1 to 8
      (1= $<$=10K, 2= $<$=15K, 3= $<$=20K, 4= $<$=25K, 5= $<$=35K
      , 6= $<$=50K, 7= $<$=75K, 8= $>$75K)

```

## B Data Exploration

### B.1 Code

The Data Exploration was carried out using the following custom code:

```

import pandas as pd
import numpy as np

# Read in the CSV data
data = pd.read_csv("diabetes_binary_health_indicators_BRFSS2015
↳ .csv")

# Open output file
with open("data_exploration.txt", "w") as file:

    # Data Size
    file.write("# Data Size\n")
    file.write(f"Number of Rows: {data.shape[0]}\n")
    file.write(f"Number of Columns: {data.shape[1]}\n\n")

```

```

# Summary Statistics
file.write("# Summary Statistics\n")
summary_stats = data.describe(include=np.number).transpose
↪ ()
summary_stats.to_csv(file, sep="\t")
file.write("\n")

# Categorical Variables
categorical_columns = data.select_dtypes(include=['object',
↪ ]).columns
file.write("# Categorical Variables\n")
if len(categorical_columns) == 0:
    file.write("Not Applicable\n\n")
else:
    for col in categorical_columns:
        file.write(f"For variable: {col}\n")
        file.write(f"Mode: {data[col].mode()[0]}\n")
    file.write("\n")

# Missing Values
file.write("# Missing Values\n")
total_missing = data.isnull().sum().sum()
if total_missing == 0:
    file.write("No missing values\n\n")
else:
    file.write(f"Total missing values: {total_missing}\n")
    missing_values = data.isnull().sum()
    missing_values.to_csv(file, sep="\t")
    file.write("\n")

```

## B.2 Code Description

The provided code performs data exploration on a given dataset of diabetes-related factors.

First, the code reads in the dataset from a CSV file and stores it in a pandas DataFrame.

The code then proceeds to analyze the dataset in several steps:

1. Data Size: The code calculates and writes the number of rows and columns in the dataset.
2. Summary Statistics: The code calculates and writes summary statistics for numerical variables in the dataset. This includes measures such as count, mean, standard deviation, minimum, and maximum for each numerical feature.
3. Categorical Variables: The code identifies and analyzes the categorical

variables in the dataset. For each categorical variable, the code writes the mode, which represents the most frequent category in the variable.

4. Missing Values: The code checks for missing values in the dataset. If there are no missing values, it writes "No missing values" in the output file. Otherwise, it calculates and writes the total number of missing values, as well as the count of missing values for each variable.

The code writes the results of the data exploration analysis to the "data\_exploration.txt" file. This includes the data size, summary statistics, categorical variables with their modes, and information about missing values.

Overall, this code provides an initial understanding of the dataset by exploring its size, summary statistics, categorical variables, and missing values. This information can be used to guide further analysis and decision-making in the research study.

### B.3 Code Output

#### data\_exploration.txt

```
\# Data Size
Number of Rows: 253680
Number of Columns: 22

\# Summary Statistics
count mean std min 25% 50% 75%
max
Diabetes\_binary 253680.0 0.1393
0.0 0.0 1.0 0.0 0.0
HighBP 253680.0 0.429 0.4949
0.0 0.0 0.0 1.0 1.0
HighChol 253680.0 0.4241 0.4942
0.0 0.0 0.0 1.0 1.0
CholCheck 253680.0 0.9627 0.1896
0.0 1.0 1.0 1.0 1.0
BMI 253680.0 28.38 6.609
12.0 24.0 27.0 31.0 98.0
Smoker 253680.0 0.4432 0.4968
0.0 0.0 0.0 1.0 1.0
Stroke 253680.0 0.04057 0.1973
0.0 0.0 0.0 0.0 1.0
HeartDiseaseorAttack 253680.0 0.09419
0.0 0.0 0.0 0.0 0.0
0.0 1.0
PhysActivity 253680.0 0.7565 0.4292
0.0 1.0 1.0 1.0 1.0
```



Fruits	253680.0	0.6343		0.4816	
		0.0	0.0	1.0	1.0
Veggies	253680.0	0.8114		0.3912	
		0.0	1.0	1.0	1.0
HvyAlcoholConsump	253680.0			0.0562	
		0.2303		0.0	0.0
	0.0	0.0	1.0		
AnyHealthcare	253680.0		0.9511		0.2158
		0.0	1.0	1.0	1.0
NoDocbcCost	253680.0		0.08418		0.2777
		0.0	0.0	0.0	1.0
GenHlth	253680.0	2.511		1.068	
		1.0	2.0	2.0	3.0
MentHlth	253680.0		3.185		7.413
		0.0	0.0	0.0	2.0
PhysHlth	253680.0		4.242		8.718
		0.0	0.0	0.0	3.0
DiffWalk	253680.0		0.1682		0.3741
		0.0	0.0	0.0	0.0
Sex	253680.0		0.4403		0.4964
		0.0	0.0	0.0	1.0
Age	253680.0		8.032		3.054
		1.0	6.0	8.0	10.0
Education	253680.0		5.05		0.9858
		1.0	4.0	5.0	6.0
Income	253680.0		6.054		2.071
		1.0	5.0	7.0	8.0

```
\# Categorical Variables
Not Applicable
```

```
\# Missing Values
No missing values
```

## C Data Analysis

### C.1 Code

The Data Analysis was carried out using the following custom code:

```
# IMPORT
import pandas as pd
import numpy as np
import statsmodels.formula.api as sm
import pickle

# LOAD DATA
```

```

data = pd.read_csv('diabetes_binary_health_indicators_BRFSS2015
    ↪ .csv')

# DATASET PREPARATIONS
# No dataset preparations are needed.

# DESCRIPTIVE STATISTICS
## Table 0: "Descriptive statistics of key variables"
key_vars = ['Diabetes_binary', 'BMI', 'PhysActivity', 'Smoker',
    ↪ 'Fruits', 'Veggies']
df0 = data[key_vars].describe().transpose() # generates
    ↪ descriptive statistics
df0 = df0[['count', 'mean', 'std']] # keeps only count, mean,
    ↪ and std
df0.to_pickle('table_0.pkl')

# PREPROCESSING
# No preprocessing is needed, because the dataset is already
    ↪ properly formatted.

# ANALYSIS
## Table 1: "Analysis of relationship between BMI and Diabetes
    ↪ moderated by Physical Activity"
model = sm.ols(formula="Diabetes_binary ~ BMI*PhysActivity +
    ↪ Age + Sex + Education + Income", data=data)
results = model.fit()
df1 = pd.DataFrame(results.summary2().tables[1]) # outputs in
    ↪ DataFrame
df1.to_pickle('table_1.pkl')

## Table 2: "Analysis of relationship between BMI and Diabetes
    ↪ moderated by Smoking"
model = sm.ols(formula="Diabetes_binary ~ BMI*Smoker + Age +
    ↪ Sex + Education + Income", data=data)
results = model.fit()
df2 = pd.DataFrame(results.summary2().tables[1]) # outputs in
    ↪ DataFrame
df2.to_pickle('table_2.pkl')

## Table 3: "Analysis of relationship between BMI and Diabetes
    ↪ moderated by Consumption of Fruits and Vegetables"
model = sm.ols(formula="Diabetes_binary ~ BMI*Fruits + BMI*
    ↪ Veggies + Age + Sex + Education + Income", data=data)
results = model.fit()
df3 = pd.DataFrame(results.summary2().tables[1]) # Outputs in
    ↪ DataFrame
df3.to_pickle('table_3.pkl')

```

```

# SAVE ADDITIONAL RESULTS
additional_results = {
    'Total number of observations': len(data),
}
with open('additional_results.pkl', 'wb') as f:
    pickle.dump(additional_results, f)

```

## C.2 Code Description

The code performs data analysis on the diabetes-related factors dataset extracted from the CDC's Behavioral Risk Factor Surveillance System (BRFSS) for the year 2015. The main goal of the analysis is to investigate the relationship between diabetes and various factors such as BMI, physical activity, smoking, and consumption of fruits and vegetables.

The analysis proceeds in several steps:

1. Dataset Loading: The code reads the dataset from the file "diabetes\_binary\_health\_indicators\_BRFSS2015.csv" and loads it into a pandas dataframe.

2. Descriptive Statistics: The code computes descriptive statistics for key variables including 'Diabetes\_binary', 'BMI', 'PhysActivity', 'Smoker', 'Fruits', and 'Veggies'. The count, mean, and standard deviation of these variables are calculated and stored in a dataframe. The dataframe is saved as "table\_0.pkl".

3. Preprocessing: Since the dataset is already clean and properly formatted, no preprocessing steps are required.

4. Analysis: The code performs three separate analyses to explore the relationship between BMI and diabetes, with the moderation of different factors:

- a. Analysis 1: The code fits a linear regression model with 'Diabetes\_binary' as the dependent variable and 'BMI', 'PhysActivity', 'Age', 'Sex', 'Education', and 'Income' as independent variables. The interaction term 'BMI\*PhysActivity' is included to examine the moderation effect of physical activity. The results, including coefficients, p-values, and confidence intervals, are stored in a dataframe. The dataframe is saved as "table\_1.pkl".

- b. Analysis 2: Similar to Analysis 1, the code fits another linear regression model with the moderation effect of smoking. The interaction term 'BMI\*Smoker' is included in the model. The results are stored in a dataframe named "table\_2.pkl".

- c. Analysis 3: The code fits a third linear regression model with the moderation effect of consuming fruits and vegetables. The interaction terms

'BMI\*Fruits' and 'BMI\*Veggies' are included in the model. The results are stored in a dataframe named "table\_3.pkl".

5. Additional Results: The code saves additional results, including the total number of observations in the dataset, in a dictionary format. The dictionary is then serialized and saved as "additional\_results.pkl" using the pickle library.

The saved outputs can be later used for further analysis, reporting, or visualization. The code provides valuable insights into the relationship between diabetes and various factors, contributing to the understanding and knowledge in the field of diabetes research.

### C.3 Code Output

**table\_0.pkl**

	count	mean	std
Diabetes\_binary	253680	0.1393	0.3463
BMI	253680	28.38	6.609
PhysActivity	253680	0.7565	0.4292
Smoker	253680	0.4432	0.4968
Fruits	253680	0.6343	0.4816
Veggies	253680	0.8114	0.3912

**table\_1.pkl**

	Coef.	Std.Err.	t	P> t
Intercept	-0.1752	0.006782	-25.83	6.23e-147
BMI	0.01198	0.000175	68.46	0
PhysActivity	0.02656	0.00647	4.104	4.06e-05
BMI:PhysActivity	-0.002213	0.0002133	-10.37	3.35e-25
Age	0.01871	0.0002163	86.48	0
Sex	0.03053	0.001329	22.98	1.08e-116
Education	-0.01118	0.0007489	-14.93	2.16e-50
Income	-0.01756	0.0003598	-48.8	0

**table\_2.pkl**

	Coef.	Std.Err.	t	P> t	[0.025	0.975]
Intercept	-0.1289	0.005873	-21.95	1.09e-106	-0.1404	-0.1174
BMI	0.009623	0.0001323	72.74	0	0.009364	0.009882
Smoker	-0.0675	0.00583	-11.58	5.39e-31	-0.07893	-0.05607
BMI:Smoker	0.002734	0.0001993	13.72	8.36e-43	0.002343	0.003125
Age	0.01899	0.0002172	87.41	0	0.01856	0.01941
Sex	0.0282	0.001336	21.1	8.91e-99	0.02558	0.03082
Education	-0.01252	0.0007488	-16.72	9.81e-63	-0.01399	-0.01105
Income	-0.01836	0.0003588	-51.17	0	-0.01907	-0.01766

table\_3.pkl

	Coef.	Std.Err.	t	P> t	[0.025	0.975]
Intercept	-0.1554	0.00797	-19.5	1.22e-84	-0.1711	-0.1398
BMI	0.0111	0.0002281	48.67	0	0.01066	0.01155
Fruits	-0.0143	0.006177	-2.315	0.0206	-0.02641	-0.002194
BMI:Fruits	0.0001439	0.0002095	0.687	0.492	-0.0002667	0.0005546
Veggies	0.004003	0.007539	0.531	0.595	-0.01077	0.01878
BMI:Veggies	-0.0005765	0.0002523	-2.285	0.0223	-0.001071	-8.191e-05
Age	0.01922	0.0002167	88.71	0	0.0188	0.01965
Sex	0.0276	0.001338	20.63	1.64e-94	0.02498	0.03022
Education	-0.01209	0.000749	-16.14	1.4e-58	-0.01356	-0.01062
Income	-0.01807	0.0003601	-50.17	0	-0.01877	-0.01736

additional\_results.pkl

```
{
  'Total number of observations': 253680,
}
```

## D LaTeX Table Design

### D.1 Code

The LaTeX Table Design was carried out using the following custom code:

```
# IMPORT
import pandas as pd
from my_utils import to_latex_with_note, is_str_in_df,
    ↪ split_mapping, AbbrToNameDef

# PREPARATION FOR ALL TABLES
shared_mapping: AbbrToNameDef = {
    'Diabetes_binary': ('Diabetes', 'Diabetes occurrence. 1 if
        ↪ yes, 0 otherwise'),
    'BMI': ('BMI', None),
    'Age': ('Age', '13-level age category in intervals of 5
        ↪ years (e.g., 1 = 18-24, 2 = 25-29)'),
    'Sex': ('Gender', '1 if male, 0 if female'),
    'Education': ('Education', 'Education Level. 1-6 with 1
        ↪ being "Never attended school" and 6 being "College
        ↪ Graduate"'),
    'Income': ('Income', 'Income Scale. 1-8 with 1 being "<=
        ↪ $10K" and 8 being ">$75K"'),
    't': ('t-val', 't-statistic of the regression estimate'),
    'P>|t|': ('p-val', 'Probability that the null hypothesis (
        ↪ of no relationship) produces results as extreme as
        ↪ the estimate')
}

# TABLE 0:
df0 = pd.read_pickle('table_0.pkl')

# DEDUPLICATE INFORMATION
count_unique = df0["count"].unique()
assert len(count_unique) == 1
df0 = df0.drop(columns=["count"])

# RENAME ROWS AND COLUMNS
mapping0 = dict((k, v) for k, v in shared_mapping.items() if
    ↪ is_str_in_df(df0, k))
mapping0['PhysActivity'] = ('Physical Activity', 'Phys.
    ↪ Activity in past 30 days, 1: Yes, 0: No')

abbrs_to_names0, legend0 = split_mapping(mapping0)
df0 = df0.rename(columns=abbrs_to_names0, index=abbrs_to_names0
    ↪ )
```

```

# SAVE AS LATEX
to_latex_with_note(
    df0, 'table_0.tex',
    caption="Descriptive statistics of key variables",
    label='table:desc_stats',
    note=f"NOTE: The number of observations in all variables is
        ↪ {count_unique[0]}",
    legend=legend0)

# TABLE 1:
df1 = pd.read_pickle('table_1.pkl')

# RENAME ROWS AND COLUMNS
mapping1 = dict((k, v) for k, v in shared_mapping.items() if
    ↪ is_str_in_df(df1, k))
mapping1['BMI:PhysActivity'] = ('BMI * Phys. Act.', '
    ↪ Interaction between BMI and Physical Activity')
mapping1['PhysActivity'] = ('Physical Activity', 'Phys.
    ↪ Activity in past 30 days, 1: Yes, 0: No')

abbrs_to_names1, legend1 = split_mapping(mapping1)
df1 = df1.rename(columns=abbrs_to_names1, index=abbrs_to_names1
    ↪ )

# SAVE AS LATEX:
to_latex_with_note(
    df1, 'table_1.tex',
    caption="Analysis of relationship between BMI and Diabetes
        ↪ moderated by Physical Activity",
    label='table:bmi_physactivity',
    legend=legend1)

# TABLE 2:
df2 = pd.read_pickle('table_2.pkl')

# RENAME ROWS AND COLUMNS
mapping2 = dict((k, v) for k, v in shared_mapping.items() if
    ↪ is_str_in_df(df2, k))
mapping2['Smoker'] = ('Smoker', '1 if smoker, 0 otherwise')
mapping2['BMI:Smoker'] = ('BMI * Smoker', 'Interaction between
    ↪ BMI and Smoking')

abbrs_to_names2, legend2 = split_mapping(mapping2)
df2 = df2.rename(columns=abbrs_to_names2, index=abbrs_to_names2
    ↪ )

# SAVE AS LATEX:
to_latex_with_note(

```

```

df2, 'table_2.tex',
caption="Analysis of relationship between BMI and Diabetes
↳ moderated by Smoking Status",
label='table:bmi_smoking',
legend=legend2)

# TABLE 3:
df3 = pd.read_pickle('table_3.pkl')

# RENAME ROWS AND COLUMNS
mapping3 = dict((k, v) for k, v in shared_mapping.items() if
↳ is_str_in_df(df3, k))
mapping3['Fruits'] = ('Fruits', 'One fruit/day, 1: Yes, 0: No')
mapping3['Veggies'] = ('Veggies', 'One veggie/day, 1: Yes, 0:
↳ No')
mapping3['BMI:Fruits'] = ('BMI * Fruits', 'Interaction between
↳ BMI and Fruit consumption')
mapping3['BMI:Veggies'] = ('BMI * Veggies', 'Interaction
↳ between BMI and Vegetable consumption')

abbrs_to_names3, legend3 = split_mapping(mapping3)
df3 = df3.rename(columns=abbrs_to_names3, index=abbrs_to_names3
↳ )

# SAVE AS LATEX:
to_latex_with_note(
    df3, 'table_3.tex',
    caption="Analysis of relationship between BMI and Diabetes
↳ moderated by Consumption of Fruits and Vegetables",
    label='table:bmi_fruits_veggies',
    legend=legend3)

```

## D.2 Provided Code

The code above is using the following provided functions:

```

def to_latex_with_note(df, filename: str, caption: str, label:
↳ str, note: str = None, legend: Dict[str, str] = None, **
↳ kwargs):
    """
    Converts a DataFrame to a LaTeX table with optional note
    ↳ and legend added below the table.

    Parameters:
    - df, filename, caption, label: as in 'df.to_latex'.
    - note (optional): Additional note below the table.

```



```

- legend (optional): Dictionary mapping abbreviations to
  ↪ full names.
- **kwargs: Additional arguments for 'df.to_latex'.
"""

def is_str_in_df(df: pd.DataFrame, s: str):
    return any(s in level for level in getattr(df.index, '
    ↪ levels', [df.index]) + getattr(df.columns, 'levels',
    ↪ [df.columns]))

AbbrToNameDef = Dict[Any, Tuple[Optional[str], Optional[str]]]

def split_mapping(abbrs_to_names_and_definitions: AbbrToNameDef
    ↪ ):
    abbrs_to_names = {abbr: name for abbr, (name, definition)
    ↪ in abbrs_to_names_and_definitions.items() if name is
    ↪ not None}
    names_to_definitions = {name or abbr: definition for abbr,
    ↪ (name, definition) in abbrs_to_names_and_definitions.
    ↪ items() if definition is not None}
    return abbrs_to_names, names_to_definitions

```

### D.3 Code Output

#### table\_0.tex

```

% This latex table was generated from: 'table\_0.pkl'
\begin{table}[h]
\caption{Descriptive statistics of key variables}
\label{table:desc\_stats}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{\%
\begin{tabular}{lrr}
\toprule
& \& mean & \& std & \& \\
\midrule
\textbf{Diabetes} & \& 0.139 & \& 0.346 & \& \\
\textbf{BMI} & \& 28.4 & \& 6.61 & \& \\
\textbf{Physical Activity} & \& 0.757 & \& 0.429 & \& \\
\textbf{Smoker} & \& 0.443 & \& 0.497 & \& \\
\textbf{Fruits} & \& 0.634 & \& 0.482 & \& \\
\textbf{Veggies} & \& 0.811 & \& 0.391 & \& \\
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize

```

```

\item NOTE: The number of observations in all variables is
253680.0
\item \textbf{Diabetes}: Diabetes occurrence. 1 if yes, 0
otherwise
\item \textbf{Physical Activity}: Phys. Activity in past 30
days, 1: Yes, 0: No
\end{tablenotes}
\end{threeparttable}
\end{table}

```

### table\_1.tex

```

% This latex table was generated from: 'table\_1.pkl'
\begin{table}[h]
\caption{Analysis of relationship between BMI and Diabetes
moderated by Physical Activity}
\label{table:bmi\_physactivity}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{\%
\begin{tabular}{lrrrlrr}
\toprule
& \& Coef. \& Std.Err. \& t-val \& p-val \& [0.025 \& 0.975] \& \\
\midrule
\textbf{Intercept} \& -0.175 \& 0.00678 \& -25.8 \& \& \& \& \$< \$1e
-06 \& -0.188 \& -0.162 \& \\
\textbf{BMI} \& 0.012 \& 0.000175 \& 68.5 \& \& \& \& \$< \$1e-06 \&
0.0116 \& 0.0123 \& \\
\textbf{Physical Activity} \& 0.0266 \& 0.00647 \& 4.1 \& \& 4.06e
-05 \& 0.0139 \& 0.0392 \& \\
\textbf{BMI * Phys. Act.} \& -0.00221 \& 0.000213 \& -10.4 \& \& \&
\& \$< \$1e-06 \& -0.00263 \& -0.00179 \& \\
\textbf{Age} \& 0.0187 \& 0.000216 \& 86.5 \& \& \& \& \$< \$1e-06 \&
0.0183 \& 0.0191 \& \\
\textbf{Gender} \& 0.0305 \& 0.00133 \& 23 \& \& \& \& \$< \$1e-06 \&
0.0279 \& 0.0331 \& \\
\textbf{Education} \& -0.0112 \& 0.000749 \& -14.9 \& \& \& \& \$< \$1e
-06 \& -0.0127 \& -0.00971 \& \\
\textbf{Income} \& -0.0176 \& 0.00036 \& -48.8 \& \& \& \& \$< \$1e-06
\& -0.0183 \& -0.0169 \& \\
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
\item \textbf{Age}: 13-level age category in intervals of 5
years (e.g., 1 = 18-24, 2 = 25-29)
\item \textbf{Gender}: 1 if male, 0 if female
\item \textbf{Education}: Education Level. 1-6 with 1 being "
Never attended school" and 6 being "College Graduate"

```

```

\item \textbf{Income}: Income Scale. 1-8 with 1 being "\$<\$
    =\$10K" and 8 being "\$>\$\$75K"
\item \textbf{t-val}: t-statistic of the regression estimate
\item \textbf{p-val}: Probability that the null hypothesis (of
    no relationship) produces results as extreme as the
    estimate
\item \textbf{BMI * Phys. Act.}: Interaction between BMI and
    Physical Activity
\item \textbf{Physical Activity}: Phys. Activity in past 30
    days, 1: Yes, 0: No
\end{tablenotes}
\end{threeparttable}
\end{table}

```

## table.2.tex

```

% This latex table was generated from: 'table\_2.pkl'
\begin{table}[h]
\caption{Analysis of relationship between BMI and Diabetes
    moderated by Smoking Status}
\label{table:bmi\_smoking}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{\%
\begin{tabular}{lrrrlrr}
\toprule
\& Coef. \& Std.Err. \& t-val \& p-val \& [0.025 \& 0.975] \& \\
\midrule
\textbf{Intercept} \& -0.129 \& 0.00587 \& -22 \& \& <\$1e-06 \\
\& -0.14 \& -0.117 \& \\
\textbf{BMI} \& 0.00962 \& 0.000132 \& 72.7 \& \& <\$1e-06 \& \\
\& 0.00936 \& 0.00988 \& \\
\textbf{Smoker} \& -0.0675 \& 0.00583 \& -11.6 \& \& <\$1e-06 \\
\& -0.0789 \& -0.0561 \& \\
\textbf{BMI * Smoker} \& 0.00273 \& 0.000199 \& 13.7 \& \& <\$ \\
\& 1e-06 \& 0.00234 \& 0.00312 \& \\
\textbf{Age} \& 0.019 \& 0.000217 \& 87.4 \& \& <\$1e-06 \& \\
\& 0.0186 \& 0.0194 \& \\
\textbf{Gender} \& 0.0282 \& 0.00134 \& 21.1 \& \& <\$1e-06 \& \\
\& 0.0256 \& 0.0308 \& \\
\textbf{Education} \& -0.0125 \& 0.000749 \& -16.7 \& \& <\$1e \\
\& -06 \& -0.014 \& -0.0111 \& \\
\textbf{Income} \& -0.0184 \& 0.000359 \& -51.2 \& \& <\$1e-06 \\
\& -0.0191 \& -0.0177 \& \\
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
\item \textbf{Age}: 13-level age category in intervals of 5

```

```

        years (e.g., 1 = 18-24, 2 = 25-29)
\item \textbf{Gender}: 1 if male, 0 if female
\item \textbf{Education}: Education Level. 1-6 with 1 being "
    Never attended school" and 6 being "College Graduate"
\item \textbf{Income}: Income Scale. 1-8 with 1 being "\$<\$
    =\$10K" and 8 being "\$>\$\$75K"
\item \textbf{t-val}: t-statistic of the regression estimate
\item \textbf{p-val}: Probability that the null hypothesis (of
    no relationship) produces results as extreme as the
    estimate
\item \textbf{Smoker}: 1 if smoker, 0 otherwise
\item \textbf{BMI * Smoker}: Interaction between BMI and
    Smoking
\end{tablenotes}
\end{threeparttable}
\end{table}

```

### table\_3.tex

```

% This latex table was generated from: 'table\_3.pkl'
\begin{table}[h]
\caption{Analysis of relationship between BMI and Diabetes
    moderated by Consumption of Fruits and Vegetables}
\label{table:bmi\_fruits\_veggies}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{\%
\begin{tabular}{lrrrrrr}
\toprule
\& Coef. \& Std.Err. \& t-val \& p-val \& [0.025 \& 0.975] \& \\
\midrule
\textbf{Intercept} \& -0.155 \& 0.00797 \& -19.5 \& \<\$1e
    -06 \& -0.171 \& -0.14 \& \\
\textbf{BMI} \& 0.0111 \& 0.000228 \& 48.7 \& \<\$1e-06 \&
    0.0107 \& 0.0116 \& \\
\textbf{Fruits} \& -0.0143 \& 0.00618 \& -2.32 \& 0.0206 \&
    -0.0264 \& -0.00219 \& \\
\textbf{BMI * Fruits} \& 0.000144 \& 0.00021 \& 0.687 \& 0.492
    \& -0.000267 \& 0.000555 \& \\
\textbf{Veggies} \& 0.004 \& 0.00754 \& 0.531 \& 0.595 \&
    -0.0108 \& 0.0188 \& \\
\textbf{BMI * Veggies} \& -0.000577 \& 0.000252 \& -2.28 \&
    0.0223 \& -0.00107 \& -8.19e-05 \& \\
\textbf{Age} \& 0.0192 \& 0.000217 \& 88.7 \& \<\$1e-06 \&
    0.0188 \& 0.0196 \& \\
\textbf{Gender} \& 0.0276 \& 0.00134 \& 20.6 \& \<\$1e-06 \&
    0.025 \& 0.0302 \& \\
\textbf{Education} \& -0.0121 \& 0.000749 \& -16.1 \& \<\$1e
    -06 \& -0.0136 \& -0.0106 \& \\
\end{tabular}
}
\end{threeparttable}
\end{table}

```

