**Table S2.** Research Goals generated by gpt-3.5-turbo and Llama 2-70B. Green highlighting indicates that a factor is a dataset feature, red highlighting indicates that this information is not a dataset feature. Results are summarized in Fig. S3A.

| | |
|---|---|
| 1<br>ChatGPT | Research Goal:<br>To investigate the association between physical activity and the likelihood of having high blood pressure, and whether this association is moderated by age.<br><br>Hypothesis:<br>There is a negative association between physical activity and the likelihood of having high blood pressure (Hypertension), and this association is moderated by age, such that the association is stronger among younger individuals compared to older individuals. |
| 2<br>ChatGPT | Research Goal:<br>To investigate the association between physical activity and the incidence of diabetes among adults in the dataset.<br><br>Hypothesis:<br>There is a negative relationship between physical activity and the risk of diabetes, such that individuals who engage in regular physical activity are less likely to have diabetes compared to those who do not engage in regular physical activity. This relationship will be moderated by age, such that the protective effect of physical activity against diabetes will be stronger among younger adults compared to older adults. |
| 3<br>ChatGPT | Research Goal:<br>To investigate the relationship between physical activity and diabetes among different age groups, while controlling for potential confounding variables. Additionally, to explore potential moderation effects of gender and education on this relationship.<br><br>Hypothesis:<br>Higher levels of physical activity are associated with a lower prevalence of diabetes, and this association is stronger among older individuals compared to younger individuals. Moreover, the association between physical activity and diabetes may be moderated by gender and education, with females and individuals with higher education experiencing a stronger protective effect of physical activity against diabetes. |

| | |
|---|---|
| 4<br>ChatGPT | Research Goal:<br>To investigate the relationship between physical activity and diabetes status, and whether this relationship is moderated by age.<br><br>Hypothesis:<br>There is a positive association between physical activity and diabetes status, such that individuals who engage in regular physical activity are less likely to have diabetes compared to those who are physically inactive. Furthermore, we hypothesize that this association will be stronger among younger individuals compared to older individuals,indicating a moderation effect of age on the relationship between physical activity and diabetes status. |
| 5<br>ChatGPT | Research Goal:<br>To examine the association between physical activity and the prevalence of diabetes, while considering the potential moderating effect of BMI.<br><br>Hypothesis:<br>There is a significant interaction effect between physical activity and BMI on the prevalence of diabetes, such that the protective effect of physical activity is stronger among individuals with lower BMI compared to those with higher BMI. |
| 6<br>ChatGPT | Research Goal:<br>To examine the relationship between fruit and vegetable consumption and the risk of diabetes, taking into account the potential interaction effect of gender.<br><br>Hypothesis:<br>Higher levels of fruit and vegetable consumption will be associated with a decreased risk of diabetes, and this association will be stronger among females compared to males. |
| 7<br>ChatGPT | Research Goal:<br>To examine the association between physical activity and the risk of diabetes, while considering the potential moderating effect of age.<br><br>Hypothesis:<br>The relationship between physical activity and the risk of diabetes will be stronger among younger individuals compared to older individuals. |

| | |
|---|---|
| 8<br>ChatGPT | Research Goal:<br>To examine the association between lifestyle factors and the presence of diabetes among individuals in the dataset.<br><br>Hypothesis:<br>There is a significant association between physical activity and the presence of diabetes, such that individuals who engage in regular physical activity are less likely to have diabetes compared to those who are physically inactive. |
| 9<br>ChatGPT | Research Goal:<br>Investigate the association between physical activity and the likelihood of having high blood pressure, and whether this association is moderated by the presence of diabetes.<br><br>Hypothesis:<br>Individuals with diabetes who engage in regular physical activity are less likely to have high blood pressure compared to individuals with diabetes who do not engage in regular physical activity. Furthermore, this association is stronger among individuals with diabetes who have a higher body mass index (BMI). |
| 10<br>ChatGPT | Research Goal:<br>To examine the association between fruit and vegetable consumption and the risk of high blood pressure, while considering the potential moderating effect of education level.<br><br>Hypothesis:<br>Higher levels of fruit and vegetable consumption will be associated with a lower risk of high blood pressure, and this association will be stronger among individuals with higher education levels. |
| 1<br>Llama | Research Goal:<br>To evaluate the effect of lower education level on glycemic control and quality of life among patients with diabetes, while considering potentially confounding factors and examining plausible mediation routes.<br><br>Hypothesis:<br>Individuals with lower education levels will exhibit inferior glycemic control and decreased quality of life relative to those with higher education levels. Furthermore, we posit that these differences will be partly attributable to variations in healthcare accessibility (including preventive care utilization, medical visit frequency, and insurance coverage) and harmful behavior patterns (principally smoking prevalence and heavy alcohol consumption). By delving into these causal pathways, the study's findings may |

| | |
|---|---|
| | contribute to the development of evidence-based strategies aimed at mitigating health disparities in disadvantaged diabetic populations. |
| 2<br>Llama | Research Goal:<br>*** Investigate the impact of social isolation on cognitive decline in older adults, taking into account potential confounding factors and identifying specific cognitive domains affected. ***<br><br>Hypothesis:<br>*** Among older adults, social isolation will be associated with accelerated cognitive decline in memory and executive functions, controlling for relevant confounders such as socioeconomic status, health conditions, lifestyle choices, urban vs. rural residency, mobility limitations, sensory impairments, and technology usage patterns. Specifically, we predict that older adults who are socially isolated, particularly those living alone and facing barriers to social engagement, will exhibit steeper declines in cognitive performance over time relative to their less isolated counterparts. Moreover, we propose that the relation between social isolation and cognitive decline will be mediated by reduced cognitive stimulation and decreased social support. **physical activity levels. Furthermore,** we expect to identify distinct cognitive profiles among older adults based on their social isolation status and explore how these profiles differ in terms of cognitive performance and decline over time. **/sprinkle throughout document** |
| 3<br>Llama | Research Goal:<br>To examine the relationship between social support and reduced symptoms of depression in individuals with diabetes, while considering the role of frequent walking or jogging.<br><br>Hypothesis:<br>Individuals with diabetes who receive social support from family, friends, and community resources will exhibit lower levels of depressive symptoms, specifically Beck Depression Inventory scores, after a six-month period, compared to those who do not receive such support. We predict that the positive effect of social support on reducing symptoms of depression will be most pronounced among those who frequently walk or jog. Furthermore, we expect that the combination of social support and frequent walking or jogging will produce a synergistic effect, leading to greater reductions in depressive symptoms compared to either intervention alone. |

| | |
|---|---|
| 4<br>Llama | Research Goal:<br>To examine the relationship between social support and mental health outcomes among adults with type 2 diabetes, and to identify potential moderators and mediators of this association.<br><br>Hypothesis:<br>Higher levels of social support will be associated with lower symptoms of depression and anxiety among adults with type 2 diabetes. We expect that this relationship will be strongest among individuals with lower levels of education, as they may face greater barriers in accessing healthcare resources and social support networks. Additionally, we will explore whether the association between social support and mental health outcomes varies by duration of diabetes diagnosis, income, gender, and comorbidities. We also test possible interaction terms between social support and these variables to uncover nuanced patterns in the data. Furthermore, we propose that increased access to emotional support, tangible aid, and informational support mediates the relationship between social support and mental health outcomes. |
| 5<br>Llama | Research Goal:<br>To investigate the relationship between social support and mental health outcomes in individuals with type 2 diabetes.<br><br>Hypothesis:<br>Individuals with type 2 diabetes who receive high levels of emotional support will have better mental health outcomes, as measured by General Health (GenHlth) and Mental Health (MentHlth), compared to those who receive low levels of emotional support. Additionally, we predict that tangible assistance will have a significant impact on MentHlth, but not GenHlth. We will also explore whether the relationship between social support and mental health outcomes varies by age, sex, and income. |
| 6<br>Llama | Research Goal:<br>To investigate the relationship between social isolation and cognitive decline in older adults with diabetes, specifically examining the role of glycemic control and potential mediating factors.<br><br>Hypothesis:<br>Older adults with diabetes who are socially isolated will experience greater cognitive decline in memory and executive function over a two-year period compared to those who are not socially isolated, after controlling for potential confounding variables. Furthermore, we predict that the association between social isolation and cognitive decline will be strongest among |

| | |
|---|---|
| | individuals with poor glycemic control, defined as HbA1c levels above 7%. We will also examine whether depression, anxiety, and sleep quality serve as mediators of the relationship between social isolation and cognitive decline. |
| 7<br>Llama | Research Goal:<br>To investigate the relationship between healthcare accessibility and health outcomes among individuals with diabetes, with a focus on understanding the temporal dynamics and potential mediating factors that influence this association.<br><br>Hypothesis:<br>We hypothesize that increased access to healthcare will lead to improved health outcomes over time among individuals with diabetes, and that this relationship will be influenced by various mediating factors, including health literacy, patient activation, and social support. Specifically, we expect that individuals with diabetes who have access to healthcare will have better HbA1c control, blood pressure management, lipid profiles, and body mass index compared to those without access, and that these differences will persist over time. Furthermore, we anticipate that the association between healthcare accessibility and health outcomes will vary across different age groups, with younger adults facing distinct barriers to healthcare access that may affect their health outcomes differently than older adults. By examining these relationships and identifying key modifiable factors,our study aims to provide actionable insights into optimizing healthcare delivery and improving health outcomes for individuals with diabetes. |
| 8<br>Llama | Research Goal:<br>=======<br>Our goal is to investigate the impact of improved cardiovascular health on healthcare utilization among patients with diabetes, specifically examining whether achieving optimal cardiovascular health metrics (e.g., normal blood pressure, cholesterol levels, and body mass index) reduces the likelihood of physician visits and, consequently, healthcare expenditure. This knowledge gap is significant since diabetes management involves long-term<br>monitoring and treatment, and understanding the effectiveness of various interventions can help policymakers allocate resources efficiently and improve patient outcomes.<br><br>Hypothesis:<br>=====<br>We hypothesize that, after adjusting for relevant confounders (age, sex, education, income, smoking |

| | |
|---|---|
| | status, and other appropriate variables), a 1 mmHg decrease in systolic blood pressure will correspond to a 2% reduction in the total number of physician visits in the past year among patients with diabetes.<br>Additionally, we expect that achieving optimal cardiovascular health markers (e.g., a combination of normal blood pressure, cholesterol levels, and body mass index) will be associated with a significantly lower likelihood of having at least one physician visit in the past year, indicating improved health outcomes and potentially reduced healthcare expenditure. This study's results can contribute to the development of evidence-based policies and clinical guidelines tailored to populations with diabetes, ultimately improving patient wellbeing while containing healthcare costs. |
| 9<br>Llama | Research Goal:<br>=======<br><br>To examine the relationship between social support from family members and mental health outcomes among individuals with diabetes, while controlling for potential confounding variables.<br><br>Hypothesis:<br>======<br><br>We hypothesize that individuals with diabetes who report higher levels of social support from family members will exhibit better mental health outcomes, including reduced symptoms of stress, anxiety, and depression, compared to those who receive less social support from family members. Furthermore, we expect that an increase in social support from family members over time will correspond with improvements in mental health outcomes. |
| 10<br>Llama | Research Goal:<br>To investigate the relationship between specific types of social support and mental health outcomes among adults with Type 2 diabetes, while controlling for potential confounders and exploring potential mediating mechanisms.<br><br>Hypothesis:<br>We hypothesize that higher levels of emotional support, practical assistance, and informational support will be associated with improved mental health outcomes (lower levels of depression, anxiety, and stress) in adults with Type 2 diabetes. Furthermore, we expect that these relationships will be moderated by factors such as age, sex, and disease duration, and that certain psychological processes (e.g., coping strategies, self-esteem) and behavioral factors (e.g., medication adherence, physical activity) will play a mediating role |

| | |
|---|---|
| | in the relationship between social support and mental health outcomes. |