Title
Estimating the Effects of Diabetes on Cardiovascular Events and Mortality: Causal Modeling and Machine Learning

Author
Inoue, Kosuke

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Estimating the Effects of Diabetes on Cardiovascular Events and Mortality: Causal Modeling and Machine Learning

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy

in Epidemiology

by

Kosuke Inoue

2021

ABSTRACT OF THE DISSERTATION

Estimating the Effects of Diabetes on Cardiovascular Events and Mortality:
Causal Modeling and Machine Learning

by

Kosuke Inoue

Doctor of Philosophy in Epidemiology
University of California, Los Angeles, 2021
Professor Beate R. Ritz, Chair

In 2020, one in ten people had diabetes in the United States. Despite the recent advancement of medical therapies, the prevalence of diabetes is still increasing, and thus more research is needed about the causal impact of diabetes and its related factors such as exercise and mental health on cardiovascular disease (CVD) and mortality. Particularly, despite the recent substantial focus on the culturally tailored and targeted approaches to improve cardiovascular health, the evidence is still limited among older Mexican Americans, a large racial/ethnic group in the US with a high prevalence of diabetes. Moreover, although elevated glycated hemoglobin (HbA1c) levels are well known to be associated with worse health outcomes, it has been under debate whether relatively lower HbA1c levels are beneficial or harmful for the long-term health outcomes

among people without diabetes. Therefore, in this dissertation, I conducted the following three studies:

First, using a longitudinal cohort of community-dwelling older Mexican Americans, along with causal mediation analysis, I found that diabetes mediated around 10% of the association of low physical activity with all-cause mortality and CVD events. Second, using the same cohort, I found that diabetes and subsequent depressive symptoms had a synergistic effect on the increased risk for cardiovascular mortality after adjusting for time-varying confounders with a marginal structural model. Third, using a nationally representative sample of US adults, along with ensemble machine learning algorithms within g-formula, I found that adults having low HbA1c levels without diabetes were at an increased risk of all-cause mortality. These findings highlight the importance of i) public health interventions targeting diabetes prevention and management among older Mexican Americans who have difficulties increasing physical activity levels, ii) mental health management for older Mexican Americans after a diagnosis of diabetes, and iii) careful monitoring of low HbA1c levels to prevent early death among US adults.

I hope that this dissertation will contribute to not only better diabetes care but a better understanding of the usefulness of causal modeling to answer clinically important questions, and will encourage dialogue and an appreciation for data sciences by clinicians and epidemiologists in the dawn of the computational era.

The dissertation of Kosuke Inoue is approved.

Elizabeth Rose Mayeda

Roch Nianogo

Donatello Telesca

Beate R. Ritz, Committee Chair

University of California, Los Angeles

2021

TABLE OF CONTENTS

CHAPTER I: Overview of this dissertation……...…….……1

CHAPTER II: Mediation of the associations of physical activity with cardiovascular events and mortality by diabetes among older Mexican Americans….....4

2.1 Introduction…….……...5

2.2 Methods…….……..……...6

2.2.1 Study Design and Participants……..……...6

2.2.2 Measurement of variables…..……..…….7

2.2.2.1 Physical Activity……..……7

2.2.2.2 Diabetes……..……..8

2.2.2.3 Other covariates……..……..8

2.2.2.4 Outcomes ascertainment……..……9

2.2.3 Statistical Analysis…..…….……10

2.3 Results..…….……..…….12

2.3.1 Demographic characteristics..……..…….12

2.3.2 Association of low physical activity with all-cause mortality, fatal CVD events, and non-fatal CVD events…..……...……...……..…...12

2.3.3 Mediation of the association of low physical activity with long-term adverse health outcomes by diabetes……..…...13

2.3.4 Stratum-specific analysis by sex……...13

2.3.4 Sensitivity analyses……..……....13

2.4 Discussion…….……..…….14

2.5 Conclusion…….……..…….18

2.6 Tables and Figures…….……..…….19

CHAPTER III: The joint association of diabetes and subsequent depressive symptoms with cardiovascular mortality among older Mexican Americans…..32
3.1 Introduction…….33
3.2 Methods…….34
3.2.1 Study Design and Participants……..34
3.2.2 Measurement of variables…..……..35
3.2.2.1 Diabetes……..35
3.2.2.2 Depressive symptoms……...35
3.2.2.3 Other covariates……..36
3.2.2.4 Cardiovascular and all-cause mortality……..36
3.2.3 Statistical Analysis…..……..37
3.3 Results..…….……..38
3.3.1 Demographic characteristics..……..38
3.3.2 Distribution of depressive symptoms across the study sample……..…...39
3.3.3 Individual association of diabetes and elevated depressive symptoms with cardiovascular and all-cause mortality ……..…...39
3.3.4 Joint association of diabetes and subsequent depressive symptoms with cardiovascular and all-cause mortality adjusting for time-varying confounders……...40
3.3.4 Sensitivity analyses...……..…….....40
3.4 Discussion…….……..41
3.5 Conclusion…….……..44
3.6 Tables and Figures..... 45

CHAPTER IV: Low HbA1c levels and all-cause or cardiovascular mortality among US adults without diabetes…….....56

4.1 Introduction…….57
4.2 Methods…….……..58
4.2.1 Study Design and Participants……..58

4.2.2 Measurement of variables…..……..…...59
4.2.2.1 Exposure and diagnosed diabetes ascertainment ….……..59
4.2.2.2 Covariates…….…….……..…...60
4.2.2.3 Outcomes ascertainment……..……..60
4.2.3 Statistical Analysis….…….……..61
4.3 Results..…….……..……63
4.3.1 Demographic characteristics..……..……63
4.3.2 Prediction of all-cause mortality and cardiovascular mortality…….…...63
4.3.3 Estimated risk of all-cause mortality and cardiovascular mortality at 5 and 10 years……..…...63
4.3.4 Stratum specific analysis by age and sex……...……...64
4.3.4 Sensitivity analyses……..…….....65
4.4 Discussion…….……..…….65
4.5 Conclusion…….……..…….69
4.6 Tables and Figures..... 70

CHAPTER V: Conclusion and Public Health Implications..……87

REFERENCES……..89

LIST OF TABLES

CHAPTER II

Table 2.1 Baseline clinical characteristics of SALSA participants according to physical activity (PA) levels 19

Table 2.2 Physical activity levels and the incidence of all-cause mortality, fatal and nonfatal CVD events 21

Table 2.3 Direct and Indirect effects (Hazard ratio scale [95%CI]) of physical activity level on the incidence of all-cause mortality, fatal and nonfatal CVD events via diabetes 22

Table S2.1 Clinical Characteristics of SALSA participants excluding those with CVD at baseline according to physical activity levels 25

Table S2.2 Competing-risks survival regression analysis for physical activity levels and incidence of fatal and non-fatal CVD events 27

Table S2.3 Sensitivity analysis additionally adjusted for metabolic factors 28

Table S2.4 Sensitivity analysis setting ≤8.3 MET-hour/week as a cut-off of low physical activity levels 29

Table S2.5 Direct and Indirect effects (Hazard ratio scale [95%CI]) of physical activity level on incidence of all-cause mortality, fatal and non-fatal CVD events via diabetes stratified by sex . 30

Table S2.6 Rate difference in additional cardiovascular events or all-cause deaths per 100,000 person-years by physical activity, separated into direct and indirect effects via diabetes..... 31

CHAPTER III

Table 3.1 Characteristics of SALSA participants according to diabetes status at enrollment..... 47

Table 3.2 Associations of (A) diabetes and (B) elevated depressive symptoms at the first follow-up visit with cardiovascular mortality and all-cause mortality 49

Table 3.3 Joint effect estimates (95% CIs) for diabetes and subsequent elevated depressive symptoms on cardiovascular mortality and all-cause mortality using marginal structural models adjusting for baseline and time-varying confounders and right censoring due to loss to follow-up at the first follow-up visit..... 50

Table S3.1 Distribution of elevated depressive symptoms at enrollment and the first follow-up visit according to diabetes status at enrollment 52

Table S3.2 Joint effect estimates (95% CIs) for diabetes and elevated depressive symptoms at enrollment with cardiovascular mortality and all-cause mortality using Cox proportional hazard models 53

Table S3.3 Joint effect estimates (95% CIs) for diabetes at enrollment and subsequent depressive symptoms at the first follow-up visit with cardiovascular mortality and all-cause mortality restricting to participants without depressive symptoms at enrollment 54

Text S3.1 Joint effect estimates (95% CIs) for diabetes at enrollment and subsequent depressive symptoms at the first follow-up visit with cardiovascular mortality and all-cause mortality restricting to participants without depressive symptoms at enrollment 55

CHAPTER IV

Table 4.1 Baseline characteristics according to glycated hemoglobin (HbA1c) levels in NHANES 1999-2014 70

Table 4.2 Predictive ability of pooled logistic regression model, tree-based algorithms, and SuperLearner for all-cause and cardiovascular mortality 72

Table 4.3 Adjusted all-cause and cardiovascular mortality risk ratio and risk difference at 5 and 10 years among participants with low glycated hemoglobin (HbA1c; 4.0 to <5.0%) compared to those with mid-level HbA1c (5.0 to <5.7%) using parametric g-formula with pooled logistic regression models 73

Table S4.1 Baseline characteristics of all covariates not described in Table 1 according to HbA1c levels in NHANES 1999-2014 78

Table S4.2 The underlying cause of death across the study population, NHANES 1999-2014 linked to mortality data through 2015 80

Table S4.3 Adjusted all-cause and cardiovascular mortality risk ratio and risk difference at 5 and 10 years among participants with prediabetes (5.7 to <6.5%) compared to those with mid-level HbA1c (5.0 to <5.7%) at 5 and 10 years using parametric g-formula with pooled logistic regression models 81

Table S4.4 Adjusted all-cause and cardiovascular mortality risk ratio and risk difference at 5 and 10 years among participants with diabetes (HbA1c≥6.5%) compared to those with mid-level HbA1c (5.0 to <5.7%) using parametric g-formula with pooled logistic regression models 82

Table S4.5 Adjusted all-cause mortality risk ratio and risk difference at 5 and 10 years among participants with low HbA1c (4.0 to <5.0%) compared to those with mid-level HbA1c (5.0 to <5.7%) stratified by age and sex using parametric g-formula with pooled logistic regression models 83

Table S4.6 Adjusted cardiovascular mortality risk ratio and risk difference at 5 and 10 years among participants with low HbA1c levels (4.0 to <5.0%) compared to those with mid-level HbA1c (5.0 to <5.7%) stratified by age and sex using parametric g-formula with pooled logistic regression models..... 84

Table S4.7 Adjusted all-cause mortality risk ratio and risk difference at 5 and 10 years among participants with low HbA1c levels (4.0 to <5.0%) compared to those with mid-level HbA1c (5.0 to <5.7%) using parametric g-formula with pooled logistic regression models, complete-case analysis (N=28,312)..... 85

Table S4.8 Adjusted all-cause mortality risk ratio and risk difference at 5 and 10 years among participants with low HbA1c levels (4.0 to <5.0%) compared to those with mid-level HbA1c (5.0 to <5.7%) using parametric g-formula with pooled logistic regression models, restricting participants to those without anemia (N=34,740)..... 86

LIST OF FIGURES

CHAPTER II

Figure 2.1 Causal diagrams illustrating causal structures under investigation 23

Figure 2.2 Decomposition of physical activity level (low vs high) on the incidence of all-cause mortality, fatal and nonfatal cardiovascular disease (CVD) events via diabetes according to sex (A: Male, B: Female): the Sacramento Area Latino Study on Aging (1998-2007)..... 24

CHAPTER III

Figure 3.1 The flow of study population, Sacramento Area Latino Study on Aging (SALSA) 1998-2007 45

Figure 3.2 Causal diagram illustrating proposed causal structure between diabetes status at enrollment, depressive symptoms at the first follow-up visit, and mortality, including time-varying metabolic disorders at enrollment and follow-up 46

Figure S3.1 Distribution of the Center for Epidemiological Studies-Depression (CESD) scores at first follow-up visit according to diabetes status at enrollment 51

CHAPTER IV

Figure 4.1 Adjusted all-cause and cardiovascular mortality risk according to glycated hemoglobin (HbA1c) levels using parametric g-formula with pooled logistic regression models and SuperLearner 74

Figure 4.2 Adjusted all-cause mortality risk among participants low glycated hemoglobin (HbA1c; 4.0 to <5.0%) and mid-level HbA1c (5.0 to <5.7%) stratified by age and sex using parametric g-formula with pooled logistic regression models..... 75

Figure S4.1 Kaplan-Meier survival curve for all-cause and cardiovascular mortality 76

Figure S4.2 Association of HbA1c with all-cause and cardiovascular mortality using restricted cubic spline curve fitted with Cox proportional hazard regression 77

ACKNOWLEDGMENTS

I would foremost like to express my deepest gratitude to my mentor and chair of the dissertation committee, Dr. Beate R. Ritz. Throughout my time as a graduate student at UCLA, she has taught me the basic concept of epidemiology, how to summarize and present the findings, and how to organize a story into an impactful scientific paper. Moreover, she has given me excellent career guidance and how to communicate with researchers. I am also deeply indebted to Dr. Elizabeth Rose Mayeda, Dr. Roch Niagnogo, and Dr. Donatello Telesca. Dr. Mayeda has been always close to me and helped to build the study design, analyze the data, and refine manuscripts and grant applications. Dr. Nianogo has supported me to understand complex causal modeling and provided helpful feedback from a statistical perspective for this dissertation. Dr. Telesca kindly introduced the world of machine learning to me, and helped me to incorporate it into causal inference framework to answer clinically important research questions. I am extremely grateful to have such a supportive dissertation committee.

I would also like to express my special appreciation to my mentors who are not members of the dissertation committee, Dr. Onyebuchi A. Arah, Dr. Zeyan Liew, and Dr. Yusuke Tsugawa. Dr. Arah has introduced me to an amazing world of causal inference, and my precious experience working with him made me decide to stay in academia as an epidemiologist focusing on the application of epidemiological methods in clinical epidemiology. Dr. Liew and Dr. Tsugawa have extensively helped me to do research and consider career development since I came to UCLA. I really enjoyed and learned a lot through intellectual discussion and exciting work with them. I also thank my best friends Ryan Assaf & Qi Yan, all of my friends at UCLA, and

collaborators around the world. Without their help, I would not have been able to spend such a wonderful time at UCLA.

Furthermore, I am grateful for the financial and academic support from the Department of Epidemiology, the Burroughs Wellcome Fund fellowship, National Institute of Health/ National Institute of Diabetes and Digestive and Kidney Diseases, Heiwa Nakajima Foundation, and Honjo International Scholarship Foundation.

Lastly, I wish to express my special thanks to my wife, Saeko Inoue. Without her dedicated support and love, I would not have been able to pursue the PhD program during such an unprecedented time of the COVID-19 pandemic. I also thank my mother Yuko Inoue, my brother Shunsuke Inoue and his family, and my grandparents Michio & Toshiko Yasuda, who have raised me and been always supportive throughout my life.

VITA

EDUCATION & TRAINING

2013	M.D. The University of Tokyo School of Medicine, Tokyo, Japan
2013-2015	Resident, Internal Medicine, National Center for Global Health and Medicine, Tokyo, Japan
2015-2017	Fellow, Endocrinology and Diabetes Center, Yokohama Rosai Hospital, Kanagawa, Japan
2018	Graduate Student Researcher, Department of Epidemiology, UCLA Fielding School of Public Health, Los Angeles, California
2019	Summer Research Intern, Harvard Global Health Institute, Boston, Massachusetts
2020	Graduate Student Researcher, Department of Radiology, David Geffen School of Medicine at UCLA, Los Angeles, California

PUBLICATIONS (*Selected*): 2017-2021, 42 publications (30 first-author)

1. **Inoue K**, Leung AM, Sugiyama T, Tsujimoto T, Makita N, Nangaku M, Ritz B. (2018). Urinary Iodine Concentrations and Mortality Among U.S. Adults. *Thyroid*; 28(7):913-920.
2. **Inoue K**, Yamazaki Y, Kitamoto T, Hirose R, Saito J, Omura M, Sasano H, Nishikawa T. (2018) Aldosterone Suppression by Dexamethasone in Patients with Aldosterone-producing Adenoma Harboring Somatic *KCNJ5* Mutations. *Journal of Clinical Endocrinology and Metabolism*; 103(9):3477-3485.
3. **Inoue K**, Ritz B, Andersen SL, Ramlau-Hansen CH, Høyer BB, Bech BH, Henriksen TB, Bonefeld-Jørgensen EC, Olsen J, Liew Z. (2019). Perfluoroalkyl Substances and maternal thyroid hormones in early pregnancy; findings in the Danish National Birth Cohort. *Environmental Health Perspectives*; 127(11):117002.
4. **Inoue K**, Nianogo R, Telesca D, Goto A, Khachadourian V, Tsugawa Y, Sugiyama T, Mayeda ER, Ritz B. (2020). Low HbA1c levels and all-cause or cardiovascular mortality among people without diabetes: the US National Health and Nutrition Examination Survey 1999-2015. *International Journal of Epidemiology* (Online ahead of print).
5. **Inoue K,** Mayeda ER, Paul K, Haan MN, Ritz B. (2020). Mediation of the Associations of Physical Activity With Cardiovascular Events and Mortality by Diabetes in Older Mexican Americans. *American Journal of Epidemiology*; 189(10):1124-1133
6. **Inoue K**, Ritz B, Ebrahimi R, Rhee C, Brent GA, Leung AM. (2020). Association of Subclinical Hypothyroidism and Cardiovascular Disease With Mortality. *JAMA Network Open*; 3(2):e1920745.

7. **Inoue K**, Yan Q, Arah OA, Paul K, Walker DI, Jones DP, Ritz B. (2020). Air pollution and adverse birth outcomes: Mediation analysis using metabolomic profiles. *Current Environmental Health Reports*; 7(3):231-242

8. **Inoue K,** Ritz B, Ernst A, Tseng WL, Yuan Y, Meng Q, Ramlau-Hansen CH, Strandberg-Larsen K, Arah OA, Obel C, Li J, Olsen J, Liew Z. (2020). Behavioral problems at age 11 years after prenatal and postnatal exposure to acetaminophen: comparing parent-reported and self-reported outcomes. *American Journal of Epidemiology* (Online ahead of print).

9. **Inoue K**, Beekely J, Goto A, Jeon C, Ritz B. (2020) Depression and Cardiovascular Disease among Patients with Type 2 Diabetes: A Systematic Review and Meta-Analysis with Bias Analysis. *Journal of Diabetes and Its Complications*; 34(12):107710.

10. **Inoue K**, Goldwater D, Allison M, Seeman T, Watson K. (2020). Serum Aldosterone Concentration, Coronary Artery Calcium, and Mortality: The Multi-Ethnic Study of Atherosclerosis. *Hypertension*; HYPERTENSIONAHA-120 -**Selected as High Impact Paper for Fall 2020 in Hypertension**-

11. **Inoue K**, Streja E, Tsujimoto T, Kobayashi H. (2020). Urinary albumin to creatinine ratio within normal range and all-cause or cardiovascular mortality among U.S. adults 1999-2015. *Annals of Epidemiology* ;55:15-23

12. **Inoue K,** Goto A, Sugiyama T, Ramlau-Hansen CH, Liew Z. (2020). The Confounder-Mediator Dilemma: Should We Control for Obesity to Estimate the Effect of Perfluoroalkyl Substances on Health Outcomes? *Toxics*; 8(4):125

13. **Inoue K**, Kitamoto T, Tsurutani Y, Saito J, Omura M, Nishikawa T. (2021). Cortisol Co-Secretion and Clinical Usefulness of ACTH Stimulation Test in Primary Aldosteronism: A Systematic Review and Biases in Epidemiological Studies. *Frontiers in Endocrinology*; 12:645488.

14. **Inoue K**, Tsugawa Y, Mangione CM, Duru OK. (2021). Association Between Industry Payments and Prescriptions of Long-acting Insulin: An Observational Study with Propensity-Score Matching. *PLOS Medicine* (in press).

15. **Inoue K**, Figueroa JF, DeJong CD, Tsugawa Y, Orav EJ, Shen C, Kazi DS. (2021). Association between Industry Marketing Payments and Prescriptions for PCSK9 inhibitors in the United States. *Circulation: Cardiovascular Quality and Outcomes* (in press).

16. Shiba K, Torres J, Daoud A, **Inoue K**, Kanamori S, Tsuji T, Kamada M, Kondo K, Kawachi I. (2021). Estimating the impact of sustained social participation on depressive symptoms in older adults. *Epidemiology* (in press).

17. Alemany S, Avella- García C, Liew Z, García-Esteban R, **Inoue K**, et al (23 authors in total). (2021). Prenatal and postnatal exposure to acetaminophen in relation to autism spectrum and attention-deficit and hyperactivity symptoms in childhood: Meta-analysis in six European population-based cohorts. *European Journal of Epidemiology* (in press).

CHAPTER I

Overview of this dissertation

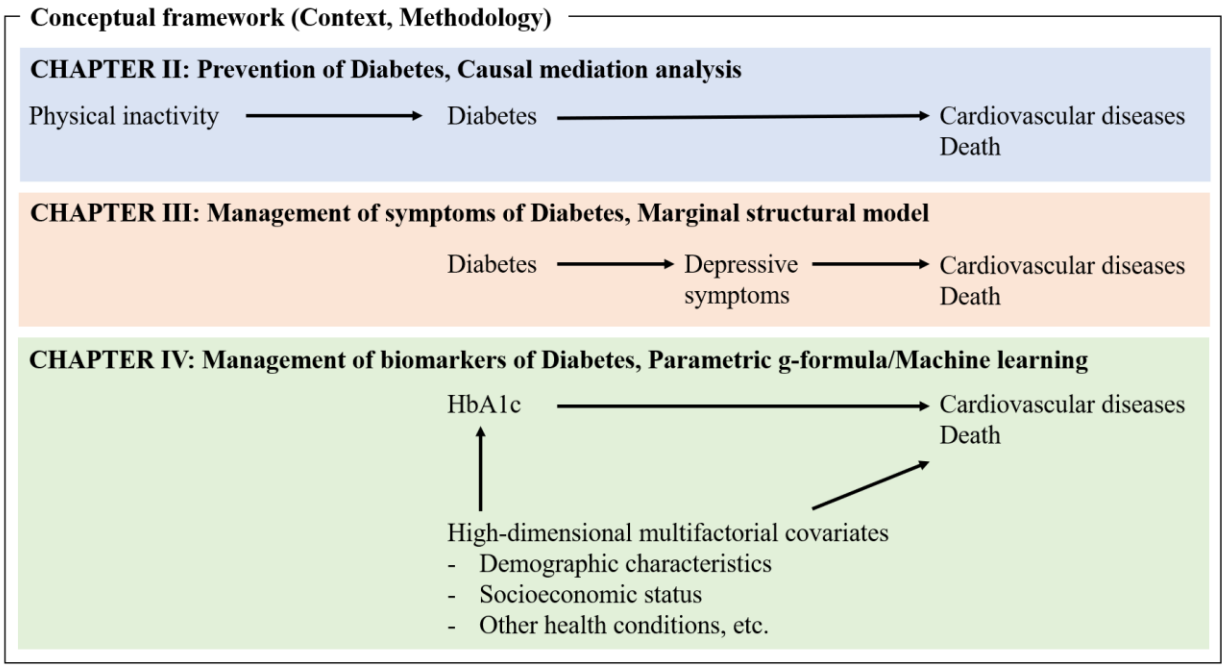
Diabetes is one of the main non-communicable diseases causing early death, affecting approximately 463 million adults in 2019, and imposes substantial health and economic burden on the global population. In response to these public health crises, as an essential part of the Sustainable Development Goals, the United Nations has proposed to reduce by one-third premature mortality from non-communicable diseases including diabetes and cardiovascular disease (CVD), by 2030. To achieve this challenging goal, it is imperative to identify the upstream and downstream factors associated with increased CVD and mortality risks due to diabetes. This is particularly important for Mexican Americans because the evidence focusing on this large race/ethnic group in the US has been very limited while they have a nearly double prevalence of type 2 diabetes as Whites. Furthermore, due to the multifactorial complex interactions between glucose metabolism and its risk factors (e.g., socio-demographic characteristics, lifestyle, comorbidities, etc), it has been challenging to address the causal pathways from glucose metabolism to long-term adverse health outcomes among the general population, which requires more studies with flexible statistical modeling on this topic.

This dissertation attempts to shed light on the following three research questions:

- To what extent does diabetes mediate the association of physical inactivity with CVD events and death among older Mexican Americans? (CHAPTER II).
- What is the influence of subsequent depressive symptoms after a diagnosis of diabetes on cardiovascular and all-cause death among older Mexican Americans? (CHAPTER III)
- Is there a causal relationship between HbA1c, a major biomarker of glucose metabolism, and death among the U.S. general population? (CHAPTER IV)

To answer these causal questions using observational studies, we applied several advanced causal inference and statistical methods including causal mediation analysis (CHAPTER II), marginal structural modeling with inverse-probability weighting (CHAPTER III), and parametric g-formula with machine learning for survival analysis (CHAPTER IV).

Figure 1.1 Conceptual framework of this dissertation



CHAPTER II

Mediation of the associations of physical activity with cardiovascular events and mortality by diabetes among older Mexican Americans

2.1 Introduction

Physical inactivity is widely recognized as an important public health problem that increases the risk of type 2 diabetes, cardiovascular disease (CVD), and mortality.[1–5] Previous studies suggest that type 2 diabetes may be a mediator on the causal pathway from physical inactivity to these long-term adverse outcomes,[6,7] but the evidence quantifying such pathway is lacking. The prevalence of physical inactivity is higher among older adults (42%) and Hispanics (40%) compared with younger adults (22-33%) and non-Hispanic whites (26%).[8] Moreover, a prior study using three national surveys showed that older Hispanics reported lower levels of physical activity (PA) than older Non-Hispanic Whites,[9] underscoring the importance of investigating the health burden of physical inactivity among this minority population.

According to the 2020 Diabetes Statistics Report, one in ten people in the United States has diabetes, and the prevalence rises to 21.4% for those aged ≥65 years.[10] While prescribing exercise is important for the management of type 2 diabetes and subsequent long-term adverse outcomes,[11] initiating and maintaining active lifestyle interventions are challenging in older adults due to multiple barriers such as comorbidities, fatigue, pain, poor perceived health, and misconceptions about benefits of PA.[12] In addition, a recent study from the National Health and Nutrition Examination Survey (NHANES) 2016-2017 reported that type 2 diabetes prevalence in Mexican Americans is nearly double compared with Non-Hispanic Whites.[10,13] The long-term health outcomes from diabetes may also differ across race/ethnicity: e.g., prior studies reported that diabetes showed weaker associations with cardiovascular disease, but stronger associations with mortality among Hispanics compared to Non-Hispanic Whites.[14,15] Therefore, understanding the causal pathway from PA to CVD and mortality through diabetes is crucial to

prevent such long-term adverse outcomes in older Mexican Americans and reduce health disparities by race/ethnicity.

In the present study, using causal mediation analysis,[16,17] we aimed to investigate whether associations between non-occupational PA and CVD (including stroke) or all-cause mortality are mediated by type 2 diabetes among older Mexican Americans. To address the potential role sex hormones may have in modifying the effects of PA on the cardiometabolic system,[18,19] we also investigated the mediation effects stratified by sex. The distinction of direct and indirect effects provides valuable information about whether public health interventions targeting diabetes prevention and management are beneficial to mitigate the overall risk of long-term adverse health outcomes among older Mexican Americans who are physically inactive.

2.2 Methods

2.2.1 Study Design and Participants

We included participants with PA information from the Sacramento Area Latino Study on Aging (SALSA), a cohort study of community-dwelling older Mexican Americans in the Sacramento area of California. Eligibility criteria included (1) being 60 years of age or older at time of enrollment in 1998–1999, (2) residing in a six-county area in the Sacramento Valley region (Sacramento, Yolo, Sutter, Solano, San Joaquin, and Placer counties), and (3) self-identification as Latino, Mexican, Central American, Mexican American. They were contacted in three stages: (i) by mail, (ii) by phone, and (iii) by door-to-door neighborhood enumeration. Participants who self-referred themselves were screened for eligibility, including residing in a targeted census tract, and household being on the sampling list. The overall response rate among those contacted

was 85%. A total of 1,789 participants were initially enrolled and they were followed with interviews and exams in their homes every 12–15 months for up to seven study visits by the end of 2007. Among the 1789 participants, 1,676 had PA information at baseline. More details about sampling and study procedures have been provided elsewhere.[20] All procedures described here were approved by the Institutional Review Boards of the University of California San Francisco, Los Angeles, and Davis, the University of Michigan, and the University of North Carolina.

2.2.2 Measurement of Variables

2.2.2.1 Physical Activity (PA)

At baseline, participants were asked to report the average number of hours they spent on 18 different types of non-work-related activities that are common among older adults. Metabolic equivalents of task (MET) were assigned to each activity based on the Compendium of PA.[21] This value was multiplied by the reported time (hours/week) spent performing the activity (MET-hours/week). Cumulative PA measures were calculated to represent moderate-vigorous intensity PA levels by summing MET-hours/week values for nine activities that require a threefold or more increase over the metabolic rate achieved by sitting quietly (≥3 METs)[22]; i.e. taking walks, walking around the neighborhood, dancing, hunting or camping or boating, swimming or engaging in workouts, golfing or other moderate exercises, gardening or yard work, house repairs, and heavy housework.[22,23] Then, participants were categorized into the following three groups according to their PA levels (MET-hours/week): low PA (<25th percentile), <20; medium PA (25th-75th percentile), 20-97; and high PA (≥75th percentile), >97.

2.2.2.2 Diabetes

We classified diabetes based on fasting glucose level ≥126 mg/dL (≥7 mmol/L), antidiabetic medication use, or self-reports of a physician diagnosis at baseline as previous studies did.[24] Fasting glucose was measured with the Cobas Mira Chemistry Analyzer (Roche Diagnostics Corporation, Indianapolis, IN). Medication use was assessed by a medicine cabinet inventory of prescription medicines.

2.2.2.3 Other Covariates

At the baseline interview, participants reported their age, gender (male, female), education levels (0, 1-8, 9-12, ≥13 years), country of birth (US or not), marital status (single, married), smoking status (current, former, never), alcohol intake levels (frequent, moderate, occasional, rarely/never), activities of daily living (ADL), instrumental activities of daily living (IADL), current working status (yes, no), and type of lifetime occupation (non-manual, manual, others). According to previous literature,[25,26] ADL limitation was defined based on whether they report difficulty in ≥1 activity or not. Similarly, IADL limitation was defined based on whether they report difficulty in ≥3 activities or not.[25,26] Acculturation was assessed using the Geriatric Acculturation Ratings Scale for Mexican Americans (G-ARSMA), a modified version of Acculturation Ratings Scale for Mexican Americans-II for use in older Latinos that consisted of 19 items assessing English and Spanish language and media use, childhood and current friendships, contact with Latin America, and dietary practices.[27] Systolic and diastolic blood pressure measurements were taken while sitting with an automatic digital blood pressure monitor and two measurements within a 10-min interval were averaged. Hypertension was based on measured systolic blood pressure (≥140 mmHg), diastolic blood pressure (≥90 mmHg), self-

report of physician diagnosis, and/or antihypertensive medication use.[28] Low-density lipoprotein (LDL) cholesterol was measured from morning fasting serum samples using the LDL Direct Liquid Select (number 7120; Equal Diagnostics). Statin prescription was also self-reported. Body mass index (BMI, kg/m^2) was calculated based on their measured height with a tape measure and weight on a Tanita scale. Waist circumference was measured at the level of maximum indentation over the abdomen with tape following a standard protocol.

2.2.2.4 Outcomes ascertainment

The primary outcome was all-cause mortality, with secondary outcomes being fatal and nonfatal CVD events (including stroke). Mortality data were ascertained through May 2010, using online obituary surveillance, a review of the Social Security Death Index and the National Death Index, a review of vital statistics data files from California, and interviews with family members. If a participant was not identified as dead, they were assumed to be alive and censored at the date of the last contact. Fatal CVD events were defined as death for which any of the following codes from the International Classification of Diseases, Tenth version (ICD–10) were mentioned anywhere on the death certificate; I20–I25, heart failure code I50, and stroke codes I63 or I64. Nonfatal CVD events were ascertained by self-report at each visit and phone call; i.e. they were asked whether a physician had diagnosed any of the following: myocardial infarction, angina, catheterization or coronary artery bypass grafting, stroke, heart failure, or atrial fibrillation. For analyses of nonfatal CVD events, 612 persons with a self-reported history of CVD at baseline were excluded to estimate the effects of PA on primary CVD events.

2.2.3 Statistical analyses

Crude and multivariable Cox proportional hazards regression models were employed for estimating effects of PA categorical exposure (low, medium, high) on all-cause mortality, fatal CVD events, and nonfatal CVD events in separate models while adjusting for potential confounders. Missing data in each variable was replaced by multiple imputations algorithms which included all of the above-mentioned covariates in the model.[29] We first adjusted for age, gender, education levels, country of birth, and marital status (Model 1). We further adjusted for G-ARSMA, smoking status, alcohol intake levels, ADL, IADL, current working status, and type of lifetime occupation in addition to Model 1 (Model 2). We also performed competing risk analysis with the method proposed by Fine and Gray considering the competing risks for fatal and nonfatal CVD events.[30] In this competing risk analysis, we estimated the subdistribution hazard by constructing risk sets that include both individuals without any event and those who have had competing events such as cancer-related mortality.[30,31]

In mediation analyses, we aimed to quantify the degree to which diabetes mediates the association between PA and long-term outcomes including all-cause mortality, fatal and nonfatal CVD events adjusting for potential confounders included in Model 2 (Figure 2.1). We employed a marginal structural approach based on the counterfactual framework to estimate the natural direct and indirect effects.[32,33] The natural direct effect is the effect of PA on long-term outcomes via pathways that do not involve diabetes while diabetes status is allowed to vary according to determinants of diabetes except PA. The natural indirect effect represents the effect of PA on long-term outcomes due to effect that PA has on diabetes; i.e. we estimate the hazard ratios of the counterfactual outcomes given a physically 'active' status if diabetes status changed to what

it would be given a physically 'inactive' status. Robust 95% confidence intervals (CIs) were estimated by repeating the analysis on 10,000 bootstrapped samples. The mediated proportion was computed as the log of the natural indirect effect divided by the log of the total effect; log(IE)/log(TE). We included cross-product terms of exposure and mediator in the model, but there was no indication of an interaction. More detailed discussion and coding tutorials using R software are described elsewhere.[16]

As previous studies have suggested that there is a difference in the effect of PA levels on long-term adverse outcomes by sex,[3,34,35] we also conducted stratum-specific analyses to estimate the causal mediation effects of diabetes on the pathway between PA and long-term adverse outcomes according to sex. P-values for the multiplicative interaction term between PA levels and sex for total effects on long-term adverse outcomes were also calculated.

We performed several sensitivity analyses. First, we additionally adjusted for other metabolic factors such as BMI, waist circumference, hypertension, statin prescription, and LDL cholesterol levels which we did not include in the main model because we assume that they are more likely to be mediators rather than confounders in the pathway from physical inactivity to CVD and mortalities. Second, we reanalyzed the data using ≤8.3 MET-hours/week (i.e. 500 MET-minute/week) as the cut-off of low physical activity levels based on the recommendation.[22,36] Finally, we fit Aalen's additive hazard models which allow us to estimate hazard differences without the assumption of proportionality.[17] Using this model, we can estimate the actual number of additional events that provide insight into the potential public health interventions. Effect estimates presented here may be considered statistically significant if the 95% confidence

interval did not include the null value. Statistical analyses were conducted using Stata version 15 and R version 3.5.2.

2.3 Results

2.3.1 Demographic characteristics

The mean age of participants was 70.3 years, and 41.7% were male (Table 2.1). PA levels were generally lower in participants with lower education levels, those with rare/never alcohol intake, but higher among those in non-manual occupations. The low PA group exhibited the highest prevalence of diabetes, hypertension, obesity, statin prescription, and history of CVD events. We found similar characteristics according to PA levels when we excluded participants with a history of CVD at baseline (Table S2.1).

2.3.2 Association of low physical activity with all-cause mortality, fatal CVD events, and non-fatal CVD events

The median duration of follow-up for all-cause mortality was 7.7 (interquartile range, 4.7–8.4) years, during which 579 deaths were identified. A total of 263 fatal CVD events and 369 nonfatal CVD events were identified. All-cause mortality was higher in the low PA group compared with the high PA group (HR, 1.36; 95% CI, 1.06–1.75) with Model 2 adjustments (Table 2.2). The low PA group also experienced higher risks of both fatal CVD (HR, 2.05; 95% CI, 1.42–2.97) and nonfatal CVD events (HR, 1.67; 95% CI, 1.18–2.37) compared with the high PA group. The medium PA group showed higher risks of nonfatal CVD compared with the high PA group (HR, 1.38; 95% CI, 1.03–1.85). The competing-risks survival regression model did not alter these results (Table 2.2, Table S2.2).

2.3.3 Mediation of the association of low physical activity with long-term adverse health outcomes by diabetes

We estimated that diabetes mediates 11.0% of the effect of PA (low vs high) on all-cause mortality (Total effect [TE], 1.36; 95% CI, 1.02-1.81 and indirect effect [IE], 1.04; 95% CI, 1.00-1.09) (Table 2.3). For fatal CVD events and nonfatal CVD events, we estimated that diabetes mediates 7.4% (TE, 2.05; 95% CI, 1.40-3.09 and IE, 1.05; 95% CI, 1.00-1.14) and 5.2% (TE, 1.67; 95% CI, 1.18-2.45 and IE, 1.03; 95% CI, 0.96-1.10), respectively. The mediation effects of diabetes on the association between PA (medium vs high) and these outcomes were small, and 95% CIs included the null. The results were qualitatively consistent when we additionally adjusted for metabolic factors (Table S2.3) and when we used the recommended physical activity levels as the cut-off point of low physical activity (Table S2.4).

2.3.4 Stratum-specific analysis by sex

In stratum-specific analyses, we estimated that diabetes mediates 55.8% of the effect of PA (low vs high) on all-cause mortality for males, but there was no evidence of mediation among females (Figure 2.2, Table S2.5). For fatal and nonfatal CVD events separately, we estimated that diabetes mediates 22.9% and 22.1% for males, but not females.

2.3.4 Sensitivity analyses

In Aalen's additive hazard model, we estimated that diabetes mediates 7.0% of the effect of PA on all-cause mortality among participants with low vs. high PA (TE, 4184 additional cases per 100,000 person-years; IE, 292 additional cases per 100,000 person-years) (Table S2.6). For fatal and nonfatal CVD events among those with low vs. high PA, we estimated that diabetes

mediates 4.3% (TE, 2744 additional cases per 100,000 person-years; IE, 118 additional cases per 100,000 person-years) and 2.2% (TE, 2929 additional cases per 100,000 person-years; IE, 63 additional cases per 100,000 person-years), respectively.

2.4 Discussion

In this population-based study of older Mexican Americans, diabetes mediated around 5-10% of the association of physical inactivity with all-cause mortality, fatal CVD events, and nonfatal CVD events. The mediation effects of diabetes on these outcomes were much more prominent among males than females.

To the best of our knowledge, this is the first study to quantify the extent to which diabetes mediates the association of PA with all-cause mortality and CVD events in older Mexican Americans. Given the high prevalence of type 2 diabetes in this population and challenges of prescribing exercise in older adults,[12,13] our findings underscore the importance of public health interventions for the prevention of type 2 diabetes and its long-term sequela among physically inactive older Mexican Americans. While it is well known that PA influences the risk of type 2 diabetes and CVD, including randomized controlled trials in older adults and meta-analyses,[6,7,37–40] to what extent type 2 diabetes mediates the association between PA and CVD has not yet been explored sufficiently. Previous studies suggesting that type 2 diabetes may mediate the association between PA and long-term adverse outcomes have approached this question by evaluating the change in estimate with and without adjusting for the potential mediators.[6,7] However, such an approach does not always validly assess mediation nor does it quantify this

effect.[17,41] Here, we used more appropriate methods i.e. proportional and additive hazard models based on the counterfactual framework.[16,32]

The underlying mechanisms through which diabetes may mediate the association between PA and long-term adverse outcomes include improved energy balance, reduction of adiposity, and reduction of inflammation through high PA.[42,43] High PA also affects myosin phenotypic characteristics, increases mitochondrial activity and volume, and increases glucose transporter type 4 protein expression, which may reduce the risk of type 2 diabetes through improved insulin sensitivity and subsequently reduce the risk of CVD and mortality.[44–46]

Consistent with prior studies,[3,34,35] the estimated total effect of low PA on long-term outcomes was larger in females than males but interaction term for PA and sex was not statistically significant. In contrast, the proportion of the estimated effect of low PA on long-term adverse outcomes mediated by diabetes was larger in males than females. The San Antonio Heart Study, a cohort study of Mexican Americans ages 25-64 years, reported an association between low PA and type 2 diabetes incidence in males only.[5] Biologically, higher PA is associated with lower levels of testosterone and estradiol in postmenopausal women but with higher testosterone levels in men.[18,19] The Mexican Americans in SALSA are older (60-93 years) than the previous studies, and PA might affect pathways involved in long-term adverse outcomes other than type 2 diabetes, e.g. endogenous levels of sex hormones, and these might have larger contributions to the overall effect of PA on adverse health outcomes in females than males. The observed sex discrepancy might also be associated with the PA health paradox; i.e. high leisure-time PA decreases risks of CVD outcomes but high occupational PA increases this risk due to sustained

inflammatory responses and 24-hour elevated heart rate.[47] Because males are more likely to have manual occupations with higher PA demands than females, the benefits of engaging in non-occupational PA for long-term adverse outcomes (especially the direct pathway that does not go through diabetes) might have been diluted by higher occupational PA levels even after controlling for current working status and type of occupation in males. Lastly, previous studies reported sex differences in type 2 diabetes due to obesity.[42,48,49] In general, females have a stronger obesity-related diabetes risk due to abdominal adiposity than males, and the impact of physical inactivity on this type of obesity is greater.[48,50,51] Thus, future studies are needed to estimate mediating effects of both obesity and type 2 diabetes from PA to long-term adverse outcomes prospectively to elucidate mechanisms underlying sex difference.[52]

The population-based longitudinal design of this study is a major strength and enabled us to study the incidence of long-term adverse outcomes over 10 years in an understudied older ethnic minority population. However, our study also has several limitations. First, while there is some evidence showing that low PA increases type 2 diabetes incidence,[34,49] and we assumed that self-reported PA reflected participants' long-term PA levels, we still have to consider the possibility of reverse causation (i.e. a diagnosis of diabetes might have affected PA levels). Those diagnosed with diabetes would be expected to be encouraged by their health care providers to increase their physical activity levels, thus, a reversal of temporality would be expected to induce a bias towards the null. Conversely, some diagnosed with diabetes a very long time ago may have become physically inactive due to complications of diabetes. Second, there is potential misclassification of the exposure, mediator, and outcomes. PA levels were estimated at baseline based on self-report, and there was no information collected regarding trends in PA levels over

the follow-up period. We classified the participants as having diabetes based on self-report, medication, and fasting glucose levels as previous studies[10,24] but lacked information on hemoglobin A1c, oral glucose tolerance test, and diabetes-related antibodies (e.g., Glutamic acid decarboxylase antibodies) at baseline. Relying on self-report of nonfatal CVD events could have introduced potential outcome misclassification. Moreover, mortality surveillance might have been less accurate after active follow-up ended in 2008. Even though misclassification was likely non-differential with respect to exposure, the bias this may generate is not always toward the null in mediation analysis.[17] Third, individuals had to survive at least to 60 years of age to participate in this study. Our estimates might have been affected by the non-enrollment of individuals with disability and mortality among people with diabetes before age 60. Forth, given that SALSA participants were residents from the Sacramento Area, our findings may not be generalizable to older Mexican Americans elsewhere. Further multi-regional studies with longitudinal measures of PA and diabetes are needed to overcome these limitations and to help better establish temporality.

Our model is based on the assumption that there are no other unmeasured confounders and no mediator-outcome confounders affected by exposure.[17] However, for example, metabolic factors (e.g., obesity, hypertension, and dyslipidemia) might be affected by PA and also affect both diabetes status and long-term health outcomes. Therefore, we may overestimate the indirect effect if we do not control for metabolic factors, and on the other hand, underestimate the direct effect when we control for metabolic factors (i.e. adjust for intermediate in the pathway not directly through diabetes). Moreover, if there is unmeasured confounding between metabolic factors and outcomes, controlling for such metabolic factors could induce collider-stratification

bias.[53] However, we found qualitatively consistent results when we adjusted for metabolic factors, indicating that these potential biases do not change our main findings substantially. As mentioned above, more advanced mediation analysis with multiple mediators would be helpful to fully address this issue in the future but require larger sample sizes.[52]

2.5 Conclusion

The present study suggests diabetes mediates the estimated effects of physical inactivity on long-term adverse outcomes among older Mexican Americans, particularly men. Given the rapidly growing older adult population with a high prevalence of diabetes and the challenges of prescribing exercise in older adults, our findings suggest that public health interventions targeting diabetes prevention and management (e.g. active diabetes screening programs focusing on older adults who are physically inactive) would be worthwhile strategies to prevent long-term adverse outcomes in older Mexican Americans.

2.6 Tables and Figures

Table 2.1 Baseline clinical characteristics of SALSA participants according to physical activity (PA) levels

Variable	Total (N=1676)		Low PA (<20 METs) (N=419)		Medium PA (20-97 METs) (N=838)		High PA (≥97 METs) (N=419)	
	N	%	N	%	N	%	N	%
Male, N (%)	699	41.7	127	30.3	356	42.5	216	51.4
Age (years old) [a]	70.3 ± 6.8		71.4 ± 7.6		70.1 ± 6.6		69.8 ± 5.9	
US born, N (%)	823	49.1	201	48.0	411	49.1	211	50.4
Education years, N (%)								
0	216	13.9	62	14.8	107	12.8	47	11.2
1-8	794	47.4	219	52.3	392	46.8	183	43.7
9-12	385	23	91	21.7	193	23.0	101	24.1
13-32 (college +)	281	16.8	47	11.2	146	17.4	88	21.0
Married, N (%)	979	58.4	223	53.2	492	58.7	264	63.0
Acculturation score [b]	22.0 ± 13.0		20.7 ± 12.7		21.9 ± 13.1		23.4 ± 12.9	
Smoking, N (%)								
Current	187	11.2	48	11.5	92	11.0	47	11.2
Former	713	42.5	168	40.1	367	43.8	178	42.5
Never	776	46.3	203	48.4	379	45.2	194	46.3
Alcohol, N (%)								
Frequent (daily)	147	8.8	27	6.5	73	8.7	47	11.2
Moderate (weekly)	181	10.8	25	6.0	89	10.6	67	16
Occasional (monthly)	153	9.1	24	5.7	86	10.3	43	10.3
Yearly/rarely/never	1195	71.3	343	81.8	590	70.4	262	62.5
ADL difficulty, [c] N (%)	200	11.9	112	26.7	66	7.9	22	5.3
IADL difficulty, [d] N(%)	788	47	266	63.6	383	45.7	139	33.2
Lifetime Occupation								
Non-manual	362	21.6	68	16.2	194	23.2	100	23.8
Manual	1000	59.7	241	57.5	496	59.2	263	62.8
Others	314	18.7	110	26.3	148	17.7	56	13.4
Currently working, N (%)	285	17	70	16.7	152	18.1	63	15.0
Statin, N (%)	141	8.4	41	9.8	68	8.1	32	7.6
LDL Cholesterol (mmol/L) [a]	3.2 ± 0.9		3.0 ± 0.9		3.2 ± 0.9		3.2 ± 0.9	
Diabetes, N (%)	542	32.3	166	39.6	256	30.6	120	28.6
Hypertension, N (%)	1136	67.8	309	73.8	547	65.3	280	66.8
BMI, kg/m^2, N (%) [e]								
<25	324	19.3	83	19.8	158	18.9	83	19.8
25-30	628	37.5	135	32.2	326	38.9	167	39.9
≥30	724	43.2	201	48.0	354	42.2	169	40.3
Waist circumference (cm) [a]	97.0 ± 13.3		98.9 ± 13.8		96.4 ± 13.4		96.2 ± 12.5	
Cardiovascular diseases, N (%)	612	36.5	187	44.6	300	35.8	125	29.8

PA, physical activity; METs, metabolic equivalents; ADL, activities of daily living; IADL, instrumental activities of daily living; LDL, low-density lipoprotein; BMI, body mass index.

[a] Values are expressed as mean (standard deviation).

[b] Acculturation was estimated using the Geriatric Acculturation Ratings Scale for Mexican Americans. The score ranges 0-56 and the lower score means less acculturated (i.e. Mexican-oriented).

[c] Participants were asked about the following difficulty to estimate ADL levels; walking, bathing, brushing hair and teeth, eating, putting clothes on, and moving from bed to chair. Difficulty of ADL was defined based on whether they report difficulty in ≥1 activity or not.

[d] Participants were asked about the following difficulty to estimate IADL: pushing objects, kneeling, lifting weights over 10 pounds, arms above shoulders, getting up from kneeling, standing up from chair, walking up stairs, writing, walking 0.25 mile, walking 10 steps, telephone, managing money, cooking, housework, and shopping. Difficulty of IADL was defined based on whether they report difficulty in ≥3 activities or not.

[e] BMI was calculated by weight (kg)/height $(m)^2$

Table 2.2 Physical activity levels and the incidence of all-cause mortality, fatal and nonfatal CVD events

Physical Activity (METs)	Incident cases	Total cases	Unadjusted		Model 1 [a]		Model 2 [b]	
			HR	95% CI	HR	95% CI	HR	95% CI
All-cause mortality								
High	115	419	1	Referent	1	Referent	1	Referent
Medium	262	838	1.21	(0.97-1.51)	1.25	(1.00-1.56)	1.13	(0.90-1.41)
Low	202	419	1.90	(1.50-2.39)	1.85	(1.46-2.34)	1.36	(1.06-1.75)
P for trend [c]			<0.001		<0.001		0.03	
Fatal CVD event								
High	46	419	1	Referent	1	Referent	1	Referent
Medium	111	838	1.25	(0.89-1.77)	1.25	(0.89-1.77)	1.07	(0.75-1.52)
Low	106	419	2.88	(2.03-4.06)	2.75	(1.93-3.91)	2.05	(1.42-2.97)
P for trend			<0.001		<0.001		0.001	
Nonfatal CVD event								
High	84	294	1	Referent	1	Referent	1	Referent
Medium	192	538	1.52	(1.15-2.02)	1.51	(1.13-2.00)	1.38	(1.03-1.85)
Low	93	232	1.97	(1.42-2.74)	1.89	(1.35-2.64)	1.67	(1.18-2.37)
P for trend			<0.001		0.001		0.004	

METs, metabolic equivalents; CVD, cardiovascular disease; ADL, activities of daily living; IADL, instrumental activities of daily living.

[a] Adjusted for age, gender, education levels, country of birth, and marital status.

[b] Adjusted for acculturation, smoking, alcohol intake levels, ADL, IADL, current working status, type of occupation in addition to Model 1.

[c] Assigned median value of METs for each physical activity category (low, 7; median, 48.5; and high, 147.5)

Table 2.3 Direct and Indirect effects (Hazard ratio scale [95%CI]) of physical activity level on the incidence of all-cause mortality, fatal and nonfatal CVD events via diabetes. [a]

	Total effect (TE)		Direct effect (DE)		Indirect effect (IE)		%mediated [b]
	HR	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**	
All-cause mortality							
Medium vs High	1.13	(0.89-1.42)	1.13	(0.87-1.42)	1.00	(0.97-1.05)	2.6
Low vs High	1.36	(1.02-1.81)	1.32	(0.99-1.77)	1.04	(1.00-1.09)	11.0
Fatal CVD events							
Medium vs High	1.07	(0.75-1.58)	1.06	(0.75-1.56)	1.00	(0.95-1.07)	6.9
Low vs High	2.05	(1.40-3.09)	1.94	(1.34-2.96)	1.05	(1.00-1.14)	7.4
Nonfatal CVD events							
Medium vs High	1.38	(1.02-1.91)	1.38	(1.03-1.92)	1.00	(0.97-1.03)	0.0
Low vs High	1.67	(1.18-2.45)	1.63	(1.17-2.41)	1.03	(0.96-1.10)	5.2

CVD, cardiovascular disease; ADL, activities of daily living; IADL, instrumental activities of daily living.

[a]Adjusted for age, gender, education levels, country of birth, marital status, acculturation, smoking, alcohol intake levels, ADL, IADL, current working status, type of occupation. Bootstrapping was performed to estimate 95% confidence interval.

[b] %mediated was calculated by log(IE)/log(TE), and therefore, depends on both total and indirect effect.

Figure 2.1 Causal diagrams illustrating causal structures under investigation.

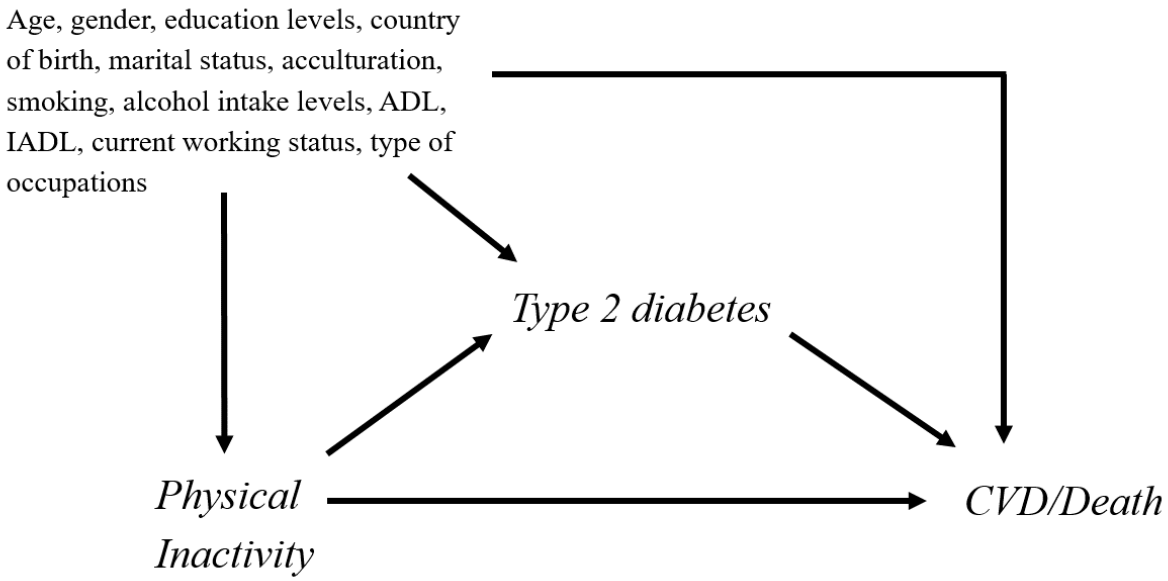
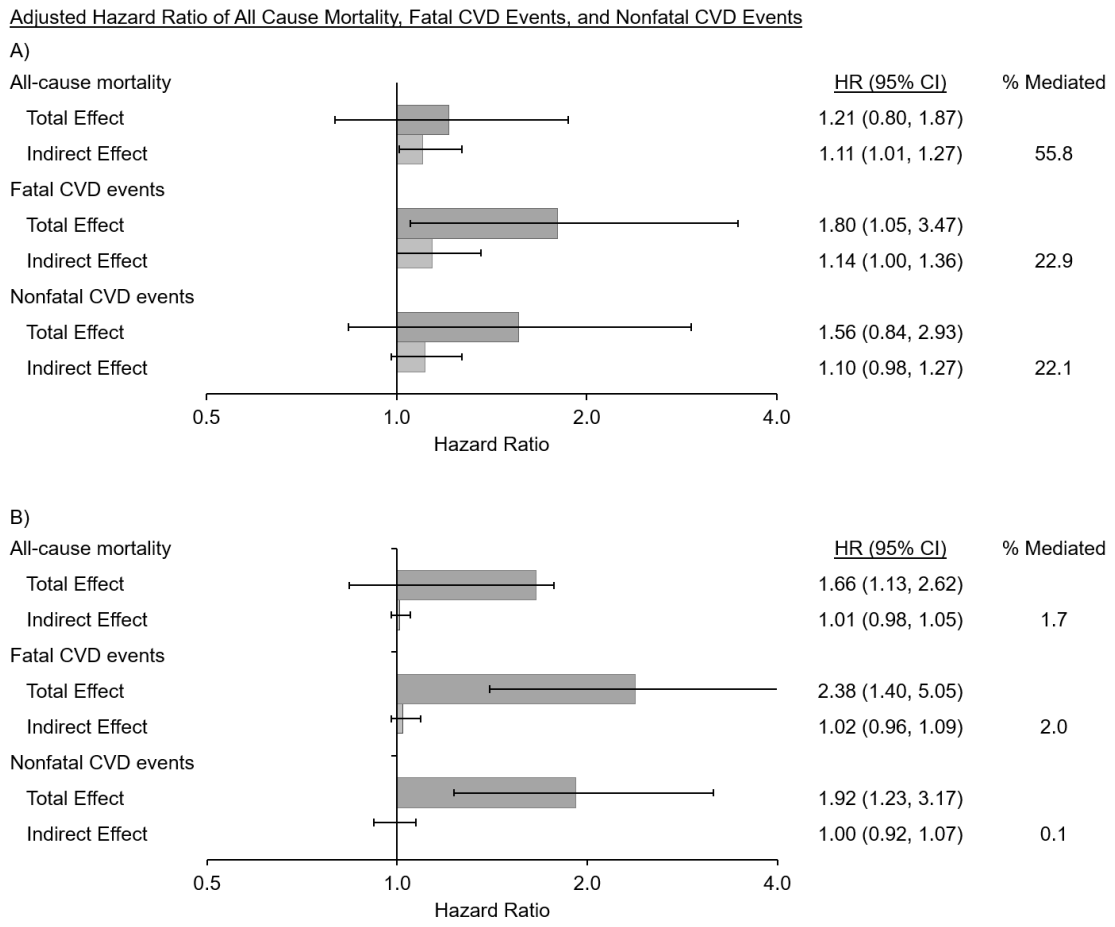


Figure 2.2 Decomposition of physical activity level (low vs high) on the incidence of all-cause mortality, fatal and nonfatal cardiovascular disease (CVD) events via diabetes according to sex (A: Male, B: Female): the Sacramento Area Latino Study on Aging (1998-2007).



Adjusted for age, gender, education levels, country of birth, marital status, acculturation, smoking, alcohol intake levels, activities of daily living (ADL), instrumental activities of daily living (IADL), current working status, type of occupation. Bootstrapping was performed to estimate 95% confidence interval (CI) of total effect (TE) and indirect effect (IE). The x-axis is shown in log-scale. %mediated was calculated by log(IE)/log(TE).
TE, total effect; IE, indirect effect (through diabetes); HR, hazard ratio

Table S2.1 Clinical Characteristics of SALSA participants excluding those with CVD at baseline according to physical activity levels

Variable	Total	Physical activity Low (<25th)	Physical activity Medium (25-75th)	Physical activity High (≥75th)
N	1064	294	538	232
METs per week		<20	20-97	≥97
Male, N (%)	464 (43.6)	78 (33.6)	234 (43.5)	152 (51.7)
Age (years old) [a]	69.9 ± 6.6	70.9 ± 7.7	69.7 ± 6.5	69.7 ± 5.9
US born, N (%)	520 (48.9)	111 (47.8)	267 (49.6)	142 (48.3)
Education years, N (%)				
0	127 (11.9)	36 (15.5)	57 (10.6)	34 (11.6)
1-8	481 (45.2)	113 (48.7)	253 (47.0)	115 (39.1)
9-12	262 (24.6)	54 (23.3)	130 (24.2)	78 (26.5)
13-	194 (18.2)	29 (12.5)	98 (18.2)	67 (22.8)
Married, N (%)	631 (59.3)	127 (54.7)	317 (58.9)	187 (63.6)
Acculturation score [b]	22.5 ± 13.0	21.4 ± 12.9	22.5 ± 13.1	23.4 ± 13.0
Smoking, N (%)				
Current	125 (11.8)	32 (13.8)	58 (10.8)	35 (11.9)
Former	421 (39.6)	83 (35.8)	225 (41.8)	113 (38.4)
Never	518 (48.7)	117 (50.4)	255 (47.4)	146 (49.7)
Alcohol, N (%)				
Frequent (daily)	104 (9.8)	19 (8.2)	51 (9.5)	34 (11.6)
Moderate (weekly)	129 (12.1)	16 (6.9)	60 (11.2)	53 (18)
Occasional (monthly)	104 (9.8)	13 (5.6)	59 (11)	32 (10.9)
Yearly/rarely/never	726 (68.3)	183 (79.2)	368 (68.4)	175 (59.5)
ADL difficulty, [c] N (%)	86 (8.1)	47 (20.3)	29 (5.4)	10 (3.4)
IADL difficulty, [d] N(%)	418 (39.3)	131 (56.5)	210 (39.0)	77 (26.2)
Lifetime occupation				
Nonmanual	234 (22.0)	35 (15.1)	125 (23.2)	74 (25.1)
Manual	648 (60.9)	140 (60.3)	320 (59.5)	188 (64.0)
Others	182 (17.1)	57 (24.6)	93 (17.3)	32 (10.9)
Currently working, N(%)	216 (20.3)	53 (22.8)	111 (20.6)	52 (17.7)
Statin, N (%)	56 (5.3)	11 (4.7)	31 (5.8)	14 (4.8)
LDL Cholesterol (mg/dL) [a]	125 ± 34	120 ± 35	126 ± 33	127 ± 35
Diabetes, N (%)	278 (26.1)	75 (32.3)	135 (25.1)	68 (23.1)
Hypertension, N (%)	660 (62.0)	158 (68.1)	322 (59.9)	180 (61.2)
BMI, kg/m^2, N (%) [e]				
<25 kg/m^2	213 (20.0)	56 (24.1)	95 (17.7)	62 (21.1)
25-30	410 (38.5)	77 (33.2)	219 (40.7)	114 (38.8)
≥30	441 (41.5)	99 (42.7)	224 (41.6)	118 (40.1)
Waist circumference (cm) [a]	96.3 ± 13.0	96.9 ± 14.0	96.3 ± 12.9	95.8 ± 12.3

ADL, activities of daily living; IADL, instrumental activities of daily living; LDL, low-density lipoprotein.

[a] mean ± standard deviation for variables with normal distribution

[b] Acculturation was estimated using the Geriatric Acculturation Ratings Scale for Mexican Americans. The score ranges 0-56 and the lower score means less acculturated (i.e. Mexican-oriented).

[c] Participants were asked about the following difficulty to estimate ADL levels; walking, bathing, brushing hair and teeth, eating, putting clothes on, and moving from bed to chair. Difficulty of ADL was defined based on whether they report difficulty in ≥1 activity or not.

[d] Participants were asked about the following difficulty to estimate IADL: pushing objects, kneeling, lifting weights over 10 pounds, arms above shoulders, getting up from kneeling, standing up from chair, walking up stairs, writing, walking 0.25 mile, walking 10 steps, telephone, managing money, cooking, housework, and shopping. Difficulty of IADL was defined based on whether they report difficulty in ≥3 activities or not.

[e] BMI was calculated by weight (kg)/height $(m)^2$

Table S2.2 Competing-risks survival regression analysis for physical activity levels and incidence of fatal and non-fatal CVD events [a]

Physical Activity (METs)	Unadjusted	Model 1 [b]	Model 2 [c]
A) Fatal CVD events			
High	Ref	Ref	Ref
Medium	1.26 (0.89-1.77)	1.27 (0.90-1.81)	1.13 (0.80-1.60)
Low	2.70 (1.91-3.81)	2.51 (1.75-3.61)	1.89 (1.29-2.76)
B) Non-fatal CVD events			
High	Ref	Ref	Ref
Medium	1.49 (1.12-1.98)	1.47 (1.10-1.95)	1.36 (1.01-1.82)
Low	1.83 (1.13-2.55)	1.74 (1.24-2.44)	1.54 (1.07-2.20)

[a] The estimated effect was interpreted as the relative change in the instantaneous *rate* of the occurrence of the event in those subjects who have not yet experienced the event of interest (but who may have experienced a competing event).
[b] Adjusted for age, gender, education levels, country of birth, and marital status.
[c] Adjusted for acculturation, smoking, alcohol intake levels, ADL, IADL, current working status, type of occupation in addition to Model 1.

Table S2.3 Sensitivity analysis additionally adjusted for metabolic factors. [a]

A) All-cause mortality	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated [b]
Medium vs High	1.15 (0.88-1.45)	1.14 (0.87-1.45)	1.01 (0.97-1.04)	3.7
Low vs High	1.38 (1.04-1.81)	1.35 (1.02-1.78)	1.02 (0.98-1.08)	7.5
B) Fatal CVD events	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated
Medium vs High	1.07 (0.75-1.60)	1.07 (0.75-1.58)	1.01 (0.96-1.07)	9.8
Low vs High	1.93 (1.32-2.92)	1.87 (1.28-2.83)	1.03 (0.98-1.10)	4.6
C) Non-fatal CVD events	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated
Medium vs High	1.39 (1.03-1.93)	1.39 (1.04-1.94)	1.00 (0.97-1.03)	0.0
Low vs High	1.60 (1.14-2.41)	1.58 (1.12-2.34)	1.01 (0.96-1.08)	2.9

[a] Adjusted for age, gender, education levels, country of birth, marital status, acculturation, smoking, alcohol intake levels, ADL, IADL, current working status, type of occupation, BMI, waist circumference at baseline, hypertension, statin prescription, and LDL cholesterol.
[b] %mediated was calculated by log(IE)/log(TE).

Table S2.4 Sensitivity analysis setting ≤8.3 MET-hour/week as a cut-off of low physical activity levels. [a]

A) All-cause mortality	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated [b]
Low vs High	1.30 (0.89-1.81)	1.28 (0.87-1.76)	1.02 (0.97-1.08)	7.2
B) Fatal CVD events	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated
Low vs High	2.13 (1.42-3.25)	2.05 (1.36-3.13)	1.04 (0.94-1.15)	5.0
C) Non-fatal CVD events	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated
Low vs High	1.90 (1.32-3.07)	1.86 (1.29-3.05)	1.02 (0.95-1.10)	3.2

[a] Adjusted for age, gender, education levels, country of birth, marital status, acculturation, smoking, alcohol intake levels, ADL, IADL, current working status, and type of occupation.
[b] %mediated was calculated by log(IE)/log(TE).

Table S2.5 Direct and Indirect effects (Hazard ratio scale [95%CI]) of physical activity level on incidence of all-cause mortality, fatal and non-fatal CVD events via diabetes stratified by sex [a,b]

Male

A) All-cause mortality	Incident cases	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated [c]
High	67/216	ref	ref	ref	-
Medium	139/356	1.08 (0.79-1.45)	1.07 (0.78-1.44)	1.01 (0.97-1.07)	15.4
Low	67/127	1.21 (0.80-1.87)	1.09 (0.72-1.69)	1.11 (1.01-1.27)	55.8
B) Fatal CVD events	Incident cases	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated
High	29/216	ref	ref	ref	-
Medium	67/356	1.13 (0.74-1.89)	1.11 (0.73-1.85)	1.02 (0.96-1.10)	14.2
Low	37/127	1.80 (1.05-3.47)	1.58 (0.94-3.06)	1.14 (1.00-1.36)	22.9
C) Non-fatal CVD events	Incident cases	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated
High	46/152	ref	ref	ref	-
Medium	87/234	1.20 (0.75-1.94)	1.20 (0.76-1.94)	1.00 (0.97-1.03)	0.2
Low	34/78	1.56 (0.84-2.93)	1.41 (0.77-2.71)	1.10 (0.98-1.27)	22.1

Female

A) All-cause mortality	Incident cases	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated
High	48/203	ref	ref	ref	-
Medium	123/482	1.21 (0.84-1.77)	1.23 (0.85-1.80)	0.99 (0.93-1.06)	NA [d]
Low	135/292	1.66 (1.13-2.62)	1.65 (1.12-2.61)	1.01 (0.98-1.05)	1.7
B) Fatal CVD events	Incident cases	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated
High	17/203	ref	ref	ref	-
Medium	44/482	1.09 (0.59-2.31)	1.10 (0.61-2.33)	0.99 (0.90-1.10)	NA [d]
Low	69/292	2.38 (1.40-5.05)	2.34 (1.37-5.02)	1.02 (0.96-1.09)	2.0
C) Non-fatal CVD events	Incident cases	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated
High	38/142	ref	ref	ref	-
Medium	105/304	1.59 (1.06-2.49)	1.60 (1.07-2.44)	0.99 (0.94-1.06)	NA [d]
Low	59/154	1.92 (1.23-3.17)	1.92 (1.25-3.23)	1.00 (0.92-1.07)	0.1

[a] Adjusted for age, gender, education levels, country of birth, marital status, acculturation, smoking, alcohol intake levels, ADL, IADL, current working status, type of occupation.

[b] P-value of multiplicative interaction between PA levels (medium vs high) and sex for total effects on all-cause mortality, fatal CVD events, and non-fatal CVD events were 0.81, 0.99, and 0.51, respectively. P-value of multiplicative interaction between PA levels (low vs high) and sex for total effects on all-cause mortality, fatal CVD events, and non-fatal CVD events were 0.14, 0.21, and 0.88, respectively.

[c] %mediated was calculated by log(IE)/log(TE).

[d] Not applicable due to the negative value of the indirect effect.

Table S2.6 Rate difference in additional cardiovascular events or all-cause deaths per 100,000 person-years by physical activity, separated into direct and indirect effects via diabetes [a]

A) All-cause mortality	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated [b]
Medium vs High	2500 (1249 to 3751)	2535 (1289 to 3781)	-35 (-102 to 33)	NA [c]
Low vs High	4184 (2566 to 5802)	3892 (2286 to 5499)	292 (207 to 376)	7.0
B) Fatal CVD events	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated
Medium vs High	1055 (218 to 1891)	1054 (227 to 1881)	1 (-8 to 10)	0.0
Low vs High	2744 (1616 to 3871)	2625 (1494 to 3756)	118 (66 to 171)	4.3
C) Non-Fatal CVD events	Total effect (TE)	Direct effect (DE)	Indirect effect (IE)	%mediated
Medium vs High	1894 (397 to 3391)	1818 (319 to 3317)	76 (16 to 136)	4.0
Low vs High	2929 (698 to 5159)	2790 (594 to 4986)	63 (-16 to 141)	2.2

[a] Adjusted for age, gender, education levels, country of birth, marital status, acculturation, smoking, alcohol intake levels, ADL, IADL, current working status, type of occupation.

[b] %mediated was calculated by log(IE)/log(TE)

[c] Not applicable due to the negative value of the indirect effect.

CHAPTER III

The joint association of diabetes and subsequent depressive symptoms with cardiovascular mortality among older Mexican Americans

3.1 Introduction

The public health burden of cardiovascular disease (CVD) and its risk factors in Hispanics have received substantial attention in the US. In 2014, the American Heart Association published a statement calling for the development of culturally tailored and targeted approaches to improve cardiovascular health and reduce CVD events among this population.[54] Diabetes is one of the major causes of CVD, affecting 14.3 million, or more than one in four US adults older than 64 years in 2018.[10] Importantly, the prevalence of diabetes varies by race/ethnicity, with nearly double the prevalence of diabetes among Hispanics compared with non-Hispanic whites.[10,13] Furthermore, diabetes is the fifth leading cause of death with 145.4 cases per 100,000 population among Hispanics ages 65 and older, while it is the seventh leading cause of death with 103.5 cases per 100,000 population among older non-Hispanic whites.[10,55] Racial/ethnic disparities in access to care and differences in adherence to treatment [14,15,56] may contribute to the higher prevalence of diabetes and its complications among Hispanics in the US. However, to design effective policy and clinical interventions that reduce racial health disparities, greater insight into factors that act as mediators and/or effect measure modifiers on the causal pathway from diabetes to CVD and death—particularly among older Hispanics, a large but understudied racial/ethnic group in the US—are needed.

Depression is a well-known factor closely associated with diabetes and CVD.[57–60] The prevalence of depression in people with type 2 diabetes is almost double that of people without diabetes.[61] In addition, a previous large case-control study revealed that after lipids and smoking, psychosocial stress, including depression, was the third factor contributing greatly to the attributable risk of acute myocardial infarction (approximately 30%).[62] A recent meta-analysis

showed that depression was associated with increased risk of CVD events among people with type 2 diabetes, and the association was robust to potential uncontrolled confounding such as employment status.[63] However, most previous studies based on conventional regression analyses have not provided sufficient insight into (i) the temporal ordering between diabetes and depression and (ii) time-varying confounders between these conditions and CVD (e.g., hypertension, obesity, and dyslipidemia). Given that depression can be both a cause or effect of diabetes[57,58,64] and they are closely interrelated with other metabolic disorders, further evidence from longitudinal studies with clear time-ordering of these diseases is warranted to evaluate the potential synergistic impact of diabetes and depression on CVD and death.

Therefore, using a longitudinal cohort of community-dwelling older Mexican Americans, along with marginal structural models (MSMs), we investigated the joint association of diabetes and subsequent depressive symptoms after a diagnosis of diabetes with cardiovascular and all-cause mortality. Fitting MSMs allows us to estimate the joint effect of two exposures at different time points (i.e. diabetes at enrollment and subsequent depressive symptoms a year after the enrollment) on outcomes (i.e. cardiovascular and all-cause mortality) in the presence of time-varying covariates that are simultaneously confounders and intermediate variables.[65]

3.2 Methods

3.2.1 Study Design and Participants

All study participants were enrolled in the Sacramento Area Latino Study on Aging (SALSA), a population-based prospective cohort of Mexican Americans aged 60 years and older. Details can be found in the previous chapter 2.2.1 or elsewhere.[20] Among 1,789 participants enrolled in

SALSA, a total of 1,495 participants (84%) with complete data on the covariates at enrollment (mentioned below) were included in this study (Figure 3.1). All procedures described here were approved by the Institutional Review Boards of the University of California San Francisco, Los Angeles, and Davis, the University of Michigan, and the University of North Carolina.

3.2.2 Measurement of Variables

3.2.2.1 Diabetes

Diabetes was classified based on fasting glucose level ≥126 mg/dL (≥7 mmol/L), antidiabetic medication use, or self-report of a physician diagnosis at enrollment using the same definition in previous studies.[24,66] Fasting glucose was measured with the Cobas Mira Chemistry Analyzer (Roche Diagnostics Corporation, Indianapolis, IN). Diabetic medications were assessed by a medicine cabinet inventory of prescription medicines.

3.2.2.2 Depressive symptoms

The Center for Epidemiological Studies-Depression (CESD), one of the most widely used measurement tools for depressive symptoms in geriatric populations[67] and older Mexican Americans,[68] was administered to all SALSA participants.[27] The CESD consists of 20 four-point (0, 1, 2, 3) Likert-type questions (total score: 0 to 60). We considered a participant to have elevated depressive symptoms when CESD was ≥16 (standard cutoff score) or they used antidepressants.[69] Antidepressant use was assessed by a medicine cabinet inventory of prescription medicines. Both CESD and antidepressant use were assessed at enrollment and the first follow-up visit (around a year after the enrollment).

3.2.2.3 Other Covariates

At baseline, participants provided sociodemographic information including age, sex, country of birth, years of education, marital status, and type of lifetime occupation (non-manual, manual, and others. We also included health behaviors related to diabetes, depression, and mortality, including smoking status, any alcohol use, and physical activity levels (METS-hour/week). Cumulative METS-hour/week was calculated by summing the self-reported average number of hours spent on nine activities with moderate-vigorous intensity as previously described.[66] Waist circumference (inches) was measured at the level of maximum indentation over the abdomen. We calculated body mass index (BMI, kg/m^2) based on measured height with a tape measure and weight on a Tanita scale. Hypertension was based on measured systolic blood pressure (≥140 mmHg), diastolic blood pressure (≥90 mmHg), self-report of physician diagnosis, and/or antihypertensives use. A previous history of CVD (including stroke) was also self-reported. Low-density lipoprotein (LDL) cholesterol was measured from morning fasting serum samples using the LDL Direct Liquid Select (Equal Diagnostics, Exton, Pennsylvania). Statin prescription was assessed using the same approach with antidiabetic medication and antidepressants. BMI, waist circumference, hypertension, previous history of CVD, and statin prescription were also assessed at the first follow-up visit.

3.2.2.4 Cardiovascular and all-cause mortality

The primary outcome was all-cause mortality, and the secondary outcome was cardiovascular mortality (including stroke). We ascertained mortality data through December 2007, using online obituary surveillance, review of the Social Security Death Index and the National Death Index, review of vital statistics data files from California, and interviews with family members. If a

participant was not identified as dead, they were assumed to be alive and censored at the date of the last contact. Cardiovascular mortality was defined based on the International Classification of Diseases, Tenth version (ICD–10); ischemic heart diseases (I20–I25), heart failure (I50), and stroke (I63-64). If a death certificate was not located, the death was coded as all-cause mortality with the unspecified cause.[70]

3.2.3 Statistical analyses

We described the distribution of CESD at the first follow-up visit according to diabetes status at enrollment. To estimate the hazard ratio (HR) of cardiovascular and all-cause mortality according to diabetes status at enrollment and elevated depressive symptoms at the first follow-up visit, we employed separate Cox proportional hazard models for diabetes and elevated depressive symptoms, respectively. We selected potential confounders at enrollment and the first follow-up visit *a priori* considering factors that may affect each outcome (i.e. cardiovascular or all-cause mortality) and might also be associated with diabetes at enrollment and elevated depressive symptoms at the first follow-up visit (Figure 3.2).

Utilizing marginal structural Cox models with inverse-probability-of-treatment weights (IPTW),[71] we investigated the association of diabetes at enrollment with cardiovascular and all-cause mortality accounting for the intermediary role of subsequent depressive symptoms at the first follow-up visit (i.e. after the assessment of diabetes status). Given the potential bias due to loss to follow-up, we also employed the inverse-probability-of-censoring weights (IPCW) assuming that diabetes status and covariates at enrollment in Figure 3.2 might have affected the censoring at the first follow-up visit. The final weights for each participant were created by

multiplying the IPTW and the IPCW (Text S3.1). Multiplicative interaction was estimated by inserting an interaction term between diabetes and subsequent depressive symptoms in the regression models, and additive interaction was estimated using the relative excess risk due to interaction (RERI). Robust 95% confidence intervals (CIs) were estimated by repeating these analyses on 1,000 bootstrapped samples.

We conducted the following two additional analyses. First, to compare the results from models with and without adjusting for time-varying metabolic disorders, we analyzed the data using diabetes status and elevated depressive symptoms at enrollment (i.e. without adjusting for time-varying metabolic disorders). Second, to minimize the possibility of reverse causation, we restricted participants without depressive symptoms at enrollment. Statistical analyses were conducted using R version 4.0.2.

3.3 Results

3.3.1 Demographic characteristics

At enrollment, the average age of participants was 70 years, and 41% were male (Table 3.1). Individuals with diabetes at enrollment generally were more likely to be born in the US, have higher BMIs and waist circumferences, have a higher prevalence of hypertension and CVD, and have a higher prevalence of statin and antidepressant use, while individuals free of diabetes at enrollment were more likely to be current smokers, report consuming alcohol, and be physically active. Among 1,495 participants included in our study, 1,136 participants (76%) completed the information on depressive symptoms and metabolic disorders at the first follow-up visit with a median duration of 1.08 (interquartile range, 0.97–1.25) years from enrollment. At the first

follow-up visit, we found similar patterns of characteristics by diabetes; i.e., individuals with diabetes at enrollment tended to have higher BMI and waist circumference, have a higher prevalence of hypertension and CVD, and higher prevalence of statin and antidepressant use at the first follow-up visit compared with individuals free of diabetes at enrollment.

3.3.2 Distribution of depressive symptoms across the study sample

The CESD at the first follow-up visit showed a right-skewed distribution, with a higher prevalence of low CESD among individuals free of diabetes than those with diabetes at enrollment (Figure S3.1). The percentage of participants with elevated depressive symptoms at the first follow-up visit was 25% (286/1136) among the total study population, 30% (107/356) among those with diabetes at enrollment, and 23% (179/780) among those without diabetes at enrollment (Table S3.1).

3.3.3 Individual association of diabetes and elevated depressive symptoms with cardiovascular and all-cause mortality

The median duration of mortality follow-up was 7.7 (interquartile range, 5.0–8.2) years. During these periods, 218 (15%) participants died from cardiovascular disease, and 341 (23%) died from all-causes in the total study population. Individuals reporting diabetes at enrollment were at increased risk of cardiovascular mortality (HR, 2.13; 95% CI, 1.60 to 2.84) and all-cause mortality (HR, 1.92; 95% CI, 1.53 to 2.41) after adjusting for potential confounders at enrollment (Table 3.2). Individuals with elevated depressive symptoms at the first follow-up visit were also at increased risk of cardiovascular mortality (HR, 1.62; 95% CI, 1.09 to 2.39) and all-

cause mortality (HR, 1.41; 95% CI, 1.02 to 1.94) after adjusting for potential confounders at enrollment and the first follow-up visit.

3.3.4 Joint association of diabetes and subsequent depressive symptoms with cardiovascular and all-cause mortality adjusting for time-varying confounders

In MSMs, individuals reporting diabetes at enrollment without depressive symptoms at the first follow-up visit were at increased risk of cardiovascular (HR, 1.82; 95% CI, 1.12 to 3.02) and all-cause mortality (HR, 2.10; 95% CI, 1.36 to 3.30) (Table 3.3). Individuals reporting diabetes at enrollment with elevated depressive symptoms at the first follow-up visit were at even greater risk of cardiovascular (HR, 5.78; 95% CI, 3.02 to 11.97) and all-cause mortality (HR, 4.32; 95% CI, 2.41 to 7.31). Multiplicative interaction between diabetes and subsequent depressive symptoms was found for cardiovascular mortality (HR, 2.94; 95% CI, 1.07 to 8.39) but not for all-cause mortality (HR, 1.80; 95% CI, 0.81 to 4.35). Additive interactions were found for both cardiovascular mortality (RERI, 3.79; 95% CI, 1.05 to 9.81) and all-cause mortality (RERI, 2.02; 95% CI, 0.01 to 5.08).

3.3.5 Sensitivity analyses

Results remained qualitatively consistent when evaluating diabetes and elevated depressive symptoms at the same time point, but we did not find interactions between diabetes and elevated depressive symptoms for cardiovascular and all-cause mortality (Table S3.2); in fact, the estimated joint effects of diabetes and elevated depressive symptoms were underestimated compared to the analysis in which we evaluated diabetes and subsequent depressive symptoms

adjusting for time-varying confounders. The results did not substantially change when we restricted the analysis to participants without depressive symptoms at enrollment (Table S3.3).

3.4 Discussion

In this longitudinal population-based study of older Mexican Americans, adjusting for time-varying metabolic disorders, we found a joint association of diabetes at baseline and subsequent elevated depressive symptoms after one year of follow-up with cardiovascular mortality. Although this observational study is not sufficient to establish causality, our findings advance our current state of knowledge and provide novel insight into the potential impact of adverse mental health after a diagnosis of diabetes on cardiovascular health outcomes.

Racial/ethnic disparities in CVD management have been one of the major public health issues in the US. Some previous studies have shown that Mexican Americans had a lower risk of cardiovascular mortality despite their higher prevalence of CVD risk factors compared with non-Hispanic whites,[72] which is sometimes discussed in the context of the "Hispanic paradox".[73,74] Diabetes has been shown to be more weakly associated with CVD but more strongly with end-stage renal disease and mortality among Mexican Americans compared to non-Hispanic whites.[14,15] Given such apparent inadequate care and paradoxes for Mexican Americans, it is imperative to better understand the forces that drive diabetes, depression, and CVD in this large racial/ethnic group living in the US with a high prevalence of diabetes.

Our findings corroborate a previously reported joint association for diabetes and depression with cardiovascular and all-cause mortality,[63] and extend the evidence to older Mexican Americans. A

prior large cohort study using a nationally representative sample of US veterans showed that comorbid depression was associated with an increased risk of CVD and all-cause mortality among people with diabetes.[75] This was also seen in another large cohort of U.S. Veterans with electronic medical records.[76] Moreover, studies have consistently reported that people affected by both diabetes and depression showed the strongest cardiovascular and all-cause mortality risks compared to people without diabetes and depression.[76–80] The findings were also true for older Mexican Americans enrolled in another cohort study (the Hispanic Established Population for the Epidemiologic Study of the Elderly survey) following its participants from 1995 through 2001.[81] However, these studies did not address the temporal order between diabetes and depression and failed to control for time-varying confounders such as metabolic disorder status. The smaller joint effects we estimated in our analysis evaluating diabetes and elevated depressive symptoms at the same time point compared to our analysis that evaluated diabetes and subsequent depressive symptoms suggest that prior findings might have underestimated the potential harmfulness of depressive symptoms among people with diabetes due to ill-defined temporality and insufficient control for confounding. In this context, our study contributes uniquely to the literature as we addressed the temporal ordering of diabetes and elevated depressive symptoms and took time-varying metabolic disorders into account.

Several mechanisms underlying the relationship between depressive symptoms and CVD have been established.[82] Depressive symptoms are known to activate the hypothalamic-pituitary-adrenal axis by releasing corticotropin-releasing factors from the hypothalamus subsequently increasing corticosteroids, which induce atherosclerosis, hypertension, and dyslipidemia.[57,82] They also decrease parasympathetic nervous system responses, which lower heart-rate variability

leading to dysrhythmia.[83] Other potential biological mechanisms that mediate the effect of depressive symptoms on CVD include inflammatory activity, endothelial dysfunction, and platelet dysfunction.[84–87] These dysfunctions may exacerbate the consequences of diabetes, metabolic syndrome, and insulin resistance, increasing the risk of having a CVD event.[88] Moreover, behavioral mechanisms play an important role in this relationship between depressive symptoms and CVD because depressive symptoms decrease adherence to medication and a healthy lifestyle such as exercise, healthy diets, and smoking cessation, all of which are protective factors for CVD.[82,89,90] These mechanisms are also strongly related to diabetes and its complications,[57] and therefore, may contribute to the synergistic effect of diabetes and depression on cardiovascular mortality.

A major strength of the present study is its population-based longitudinal design and follow-up of older Mexican Americans for about a decade that allowed us to investigate the incidence of cardiovascular and all-cause mortality in an understudied racial/ethnic group. Moreover, this resource helped us clarify the temporal relationship between diabetes and elevated depressive symptoms while at the same time adjusting for time-varying metabolic disorders employing IPTW. We further utilized IPCW to account for potential bias due to loss to follow-up. However, our study has several limitations. First, our findings may not be generalizable to institutionalized older Mexican Americans or those living outside the Sacramento Area. Second, cohort participants had to survive to at least 60 years of age to be enrolled, and therefore, our results might have been biased due to the exclusion of people who died before age 60 years. Third, although we adjusted for an extensive set of potential confounders, there is a potential for uncontrolled or residual confounding due to the nature of the observational design. Lastly,

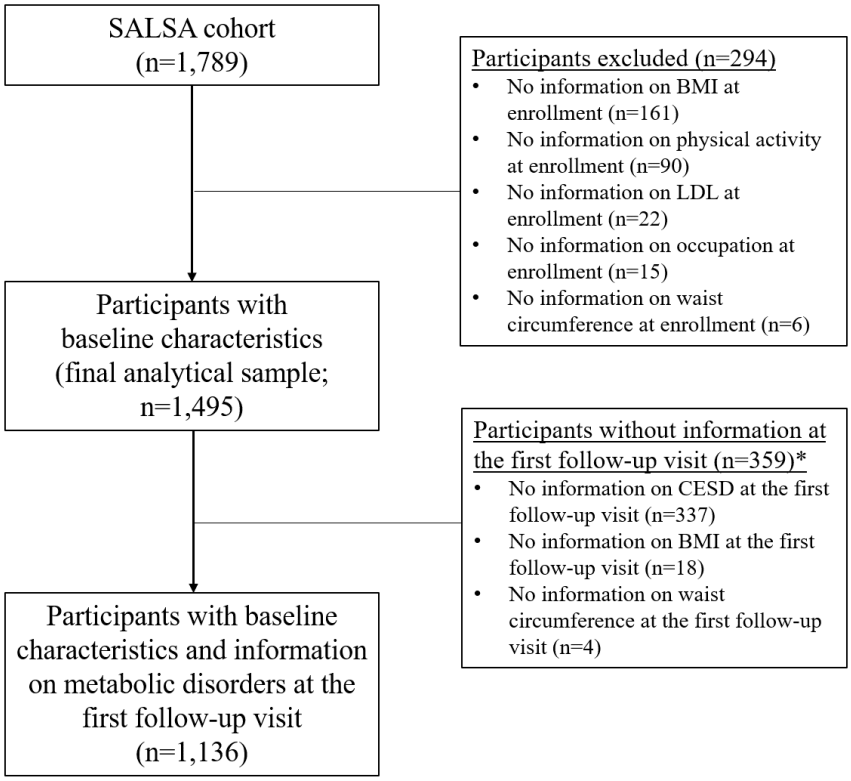
because we classified diabetes based on self-report, medication, and fasting glucose levels, we cannot rule out the possibility of misclassification of this exposure. We did not have information about diabetes severity, duration of diabetes, and whether depressive symptoms and mortality were directly related to diabetes.

3.5 Conclusion

Using the longitudinal cohort with clear temporal ordering between diabetes and elevated depressive symptoms, along with employing methods to adjust for time-varying metabolic disorders, we found that diabetes and subsequent depressive symptoms were jointly associated with an increased risk of cardiovascular mortality among community-dwelling older Mexican Americans. Future studies are needed to illuminate whether and what kind of clinical interventions to reduce depressive symptoms after diabetes are beneficial to promote cardiovascular health among older Mexican Americans.

3.6 Tables and Figures

Figure 3.1 The flow of study population, Sacramento Area Latino Study on Aging (SALSA) 1998-2007.



*Inverse-probability-censoring weights were applied to adjust for the right censoring at the first follow-up visit due to loss to follow-up (n=359).

Figure 3.2 Causal diagram illustrating proposed causal structure between diabetes status at enrollment, depressive symptoms at the first follow-up visit, and mortality, including time-varying metabolic disorders at enrollment and follow-up.

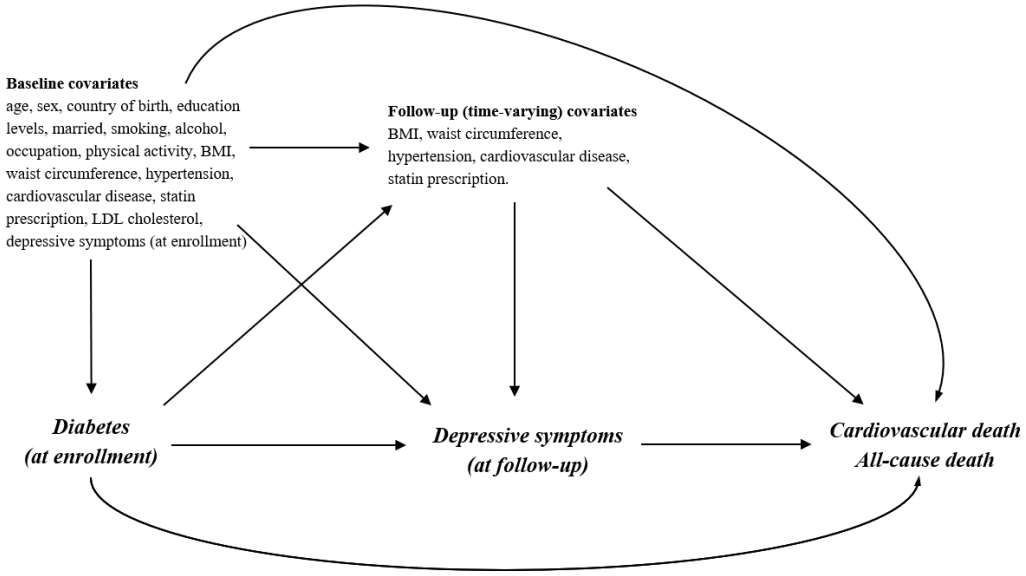


Table 3.1 Characteristics of SALSA participants according to diabetes status at enrollment.[a]

Variable	Diabetes at enrollment (N=488)	Free of diabetes at enrollment (N=1007)
A) Baseline information		
Age (years old)	69.9 ± 6.5	70.2 ± 6.7
Male, N (%)	213 (43.7)	405 (40.2)
US born, N (%)	280 (57.4)	472 (46.9)
Education years, N (%)		
0	68 (13.9)	117 (11.6)
1-8	220 (45.1)	478 (47.5)
9-12	117 (24.0)	234 (23.2)
≥13	83 (17.0)	178 (17.7)
Married, N (%)	295 (60.5)	576 (57.2)
Type of occupation		
Non-manual	109 (22.3)	220 (21.9)
Manual	287 (58.8)	596 (59.2)
Other	92 (18.9)	191 (19.0)
Smoking, N (%)		
Current	45 (9.2)	126 (12.5)
Former	229 (46.9)	412 (40.9)
Never	214 (43.9)	469 (46.6)
Alcohol consumption, N (%)	203 (41.6)	622 (61.8)
Physical Activity (METs per week)	64.3 ± 72.5	75.2 ± 74.4
BMI (kg/m^2)	31.1 ± 6.3	29.2 ± 5.6
Waist circumference (inches)	39.9 ± 4.8	37.4 ± 5.2
Hypertension, N (%)	400 (82.0)	637 (63.3)
Cardiovascular diseases, N (%)	234 (48.0)	308 (30.6)
LDL Cholesterol (mg/dL)	117 ± 35	126 ± 34
Statin use, N (%)	54 (11.1)	82 (8.1)
CESD scale	10.7 ± 10.7	9.5 ± 10.4
Anti-depressant use, N(%)	55 (11.3)	69 (6.9)
B) Follow-up information (at the first follow-up visit)	**Diabetes at enrollment (N=356)**	**Free of diabetes at enrollment (N=780)**
Years from enrollment to the first follow-up	1.12 ± 0.21	1.13 ± 0.21
BMI (kg/m^2)	31.0 ± 6.5	29.2 ± 5.8
Waist circumference (inches)	39.6 ± 5.2	37.6 ± 5.2

Hypertension, N (%)	310 (87.1)	574 (73.6)
Cardiovascular diseases, N (%)	198 (55.6)	270 (34.6)
Statin, N (%)	62 (17.4)	103 (13.2)
CESD scale	9.7 ± 10.4	7.9 ± 9.6
Anti-depressant use, N(%)	41 (11.5)	63 (8.1)

SALSA, Sacramento Area Latino Study on Aging; METs, metabolic equivalent for task; BMI, body mass index; LDL, low-density lipoprotein; CESD, the Center for Epidemiological Studies-Depression.
[a]Data are presented as count (percentage) or mean ± standard deviation otherwise indicated

Table 3.2 Associations of (A) diabetes and (B) elevated depressive symptoms at the first follow-up visit with cardiovascular mortality and all-cause mortality.

Outcomes	Cardiovascular mortality		All-cause mortality	
Exposures	**Number of Events**	**Adjusted HR (95% CI)**	**Number of Events**	**Adjusted HR (95% CI)**
A) Diabetes at enrollment [a]				
No	102/1007	Ref	178/1007	Ref
Yes	116/488	2.13 (1.60 to 2.84)	163/488	1.92 (1.53 to 2.41)
B) Elevated depressive symptoms at the first follow-up visit [b]				
No	93/850	Ref	150/850	Ref
Yes	55/286	1.62 (1.09 to 2.39)	81/286	1.41 (1.02 to 1.94)

HR, hazard ratio; CI, confidence interval; BMI, body mass index; LDL, *Low*-Density Lipoprotein cholesterol.

[a]Adjusted for baseline covariates (age, sex, country of birth, education levels, marital status, type of occupation, physical activity, smoking status, alcohol drinking, physical activity, BMI, waist circumference, hypertension, cardiovascular diseases, LDL cholesterol, statin prescription).

[b]Adjusted for diabetes status at enrollment, baseline covariates (age, sex, country of birth, education levels, marital status, type of occupation, physical activity, smoking status, alcohol drinking, physical activity, BMI, waist circumference, hypertension, cardiovascular diseases, LDL cholesterol, statin prescription, elevated depressive symptoms at enrollment), and time-varying covariates (BMI, waist circumference, hypertension, cardiovascular diseases, statin prescription) at the first follow-up visit.

Table 3.3 Joint effect estimates (95% CIs) for diabetes and subsequent elevated depressive symptoms on cardiovascular mortality and all-cause mortality using marginal structural models adjusting for baseline and time-varying confounders and right censoring due to loss to follow-up at the first follow-up visit.

Outcomes		**Cardiovascular mortality**		**All-cause mortality**	
Diabetes at enrollment	**Elevated depressive symptoms at follow-up**	**Number of Events**	**Adjusted HR (95% CI)** [a, b]	**Number of Events**	**Adjusted HR (95% CI)** [a, b]
No	No	50/601	Ref	86/601	Ref
Yes	No	43/249	1.82 (1.12 to 3.02)	64/249	2.10 (1.36 to 3.30)
No	Yes	21/179	1.09 (0.46 to 2.17)	38/179	1.13 (0.65 to 1.88)
Yes	Yes	34/107	5.78 (3.02 to 11.97)	43/107	4.32 (2.41 to 7.31)
HR for the interaction term (multiplicative scale) [c]			2.94 (1.07 to 8.39)		1.80 (0.81 to 4.35)
RERI (additive scale) [c]			3.79 (1.05 to 9.81)		2.02 (0.01 to 5.08)

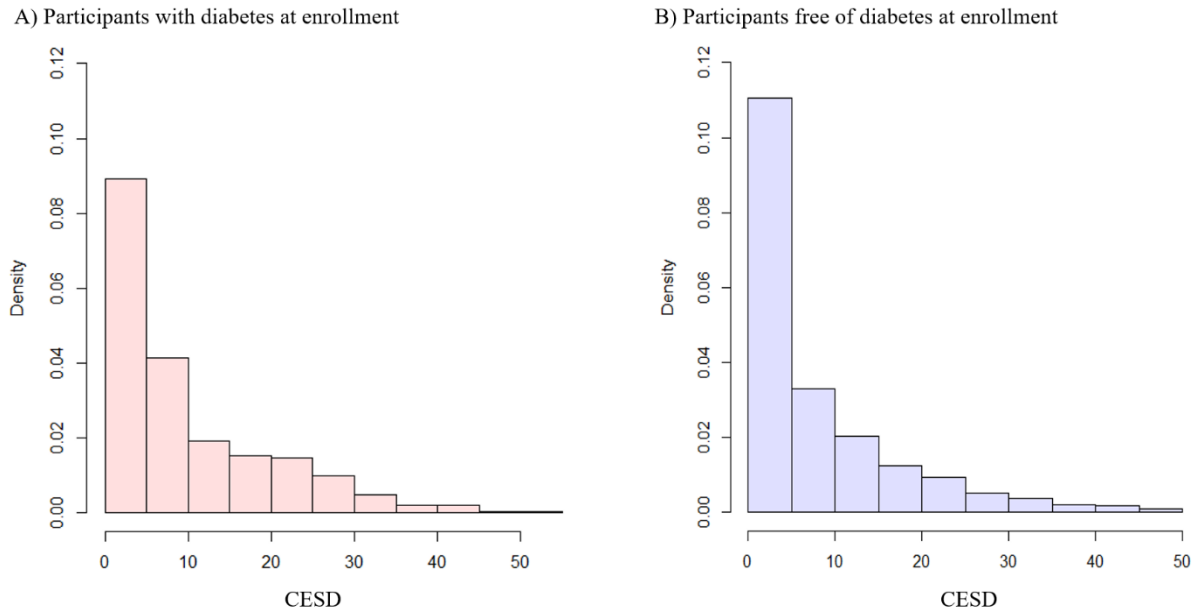
HR, hazard ratio; CI, confidence interval; RERI, relative excess risk due to interaction; BMI, body mass index; LDL, Low-Density Lipoprotein cholesterol.

[a] Inverse probability of treatment weights was applied to adjust for baseline covariates (age, sex, country of birth, education levels, marital status, type of occupation, physical activity, smoking status, alcohol drinking, physical activity, BMI, waist circumference, hypertension, cardiovascular diseases, LDL cholesterol, statin prescription, elevated depressive symptoms at enrollment) and covariates at the first follow-up visit (BMI, waist circumference, hypertension, cardiovascular diseases, statin prescription). Inverse probability of censoring weights was also applied to adjust for right censoring at the first follow-up visit due to loss to follow-up.

[b] 1000 iterations were performed for bootstrapping to estimate 95% CI.

[c] The interaction was significant for cardiovascular mortality on both multiplicative and additive scales, and was significant for all-cause mortality on the additive scale. Multiplicative interaction was calculated by inserting multiplicative term between diabetes and elevated depressive symptoms ($HR_{DM(yes)\_Dep(yes)}$ /[$HR_{DM(yes)\_Dep(no)} \times HR_{DM(no)\_Dep(yes)}$]; null value = 1), and additive interaction was calculated by RERI ($HR_{DM(yes)\_Dep(yes)}$ - $HR_{DM(yes)\_Dep(no)}$ - $HR_{DM(no)\_Dep(yes)}$ +1; null value = 0).

Figure S3.1 Distribution of the Center for Epidemiological Studies-Depression (CESD) scores at first follow-up visit according to diabetes status at enrollment



The percentage of reporting anti-depressant use was 11.3% and 6.9% among participants with diabetes at enrollment and those free of diabetes at enrollment, respectively.

Table S3.1 Distribution of elevated depressive symptoms at enrollment and the first follow-up visit according to diabetes status at enrollment.

A) Total population		Elevated depressive symptoms at follow-up		
		Yes	No	Total
Elevated depressive symptoms at enrollment	Yes	177	148	325
	No	109	702	811
	Total	286	850	1136
B) Participants with diabetes at enrollment		Elevated depressive symptoms at follow-up		
		Yes	No	Total
Elevated depressive symptoms at enrollment	Yes	66	53	119
	No	41	196	237
	Total	107	249	356
C) Participants free of diabetes at enrollment		Elevated depressive symptoms at follow-up		
		Yes	No	Total
Elevated depressive symptoms at enrollment	Yes	111	95	206
	No	68	506	574
	Total	179	601	780

Table S3.2 Joint effect estimates (95% CIs) for diabetes and elevated depressive symptoms at enrollment with cardiovascular mortality and all-cause mortality using Cox proportional hazard models.

Outcomes		**Cardiovascular mortality**		**All-cause mortality**	
Diabetes at enrollment	**Elevated depressive symptoms at enrollment**	**Number of Events**	**Adjusted HR (95% CI)** [a, b]	**Number of Events**	**Adjusted HR (95% CI)** [a b]
No	No	62/724	Ref	111/724	Ref
Yes	No	78/318	2.56 (1.75 to 3.65)	106/318	2.12 (1.59 to 2.87)
No	Yes	40/283	1.38 (0.86 to 2.18)	67/283	1.29 (0.94 to 1.79)
Yes	Yes	38/170	2.22 (1.39 to 2.18)	57/170	2.12 (1.49 to 3.16)
HR for the interaction term (multiplicative scale) [c]			0.64 (0.35 to 1.18)		0.78 (0.48 to 1.27)
RERI (additive scale) [c]			-0.68 (-2.00 to 0.48)		-0.26 (-1.23 to 0.59)

HR, hazard ratio; CI, confidence interval; RERI, relative excess risk due to interaction; BMI, body mass index; LDL, *Low*-Density Lipoprotein cholesterol.

[a]Adjusted for baseline covariates (age, sex, country of birth, education levels, marital status, type of occupation, physical activity, smoking status, alcohol drinking, physical activity, BMI, waist circumference, hypertension, cardiovascular diseases, LDL cholesterol, statin prescription).

[b] 1000 iterations were performed for bootstrapping to estimate 95% CI.

[c] The interaction was significant for cardiovascular mortality on both multiplicative and additive scales, and was significant for all-cause mortality on the additive scale. Multiplicative interaction was calculated by inserting multiplicative term between diabetes and elevated depressive symptoms ($HR_{DM(yes)\_Dep(yes)}$ /[$HR_{DM(yes)\_Dep(no)} \times HR_{DM(no)\_Dep(yes)}$]; null value = 1), and additive interaction was calculated by RERI ($HR_{DM(yes)\_Dep(yes)}$ - $HR_{DM(yes)\_Dep(no)}$ - $HR_{DM(no)\_Dep(yes)}$ +1; null value = 0).

Table S3.3 Joint effect estimates (95% CIs) for diabetes at enrollment and subsequent depressive symptoms at the first follow-up visit with cardiovascular mortality and all-cause mortality restricting participants without depressive symptoms at enrollment.

Outcomes		**Cardiovascular mortality**		**All-cause mortality**	
Diabetes at enrollment	**Elevated depressive symptoms at follow-up**	**Number of Events**	**Adjusted HR (95% CI)** [a]	**Number of Events**	**Adjusted HR (95% CI)** [a]
No	No	38/506	Ref	68/506	Ref
Yes	No	38/196	2.40 (1.34 to 4.35)	53/196	2.19 (1.47 to 3.48)
No	Yes	6/68	0.75 (0.12 to 2.11)	12/68	0.93 (0.33 to 1.98)
Yes	Yes	15/41	7.12 (2.86 to 18.06)	19/41	5.23 (2.56 to 10.51)
HR for the interaction term (multiplicative scale) [c]			4.28 (0.96 to 28.90)		2.58 (0.92 to 8.66)
RERI (additive scale) [c]			4.93 (-0.48 to 14.75)		2.98 (0.17 to 8.43)

HR, hazard ratio; CI, confidence interval; RERI, relative excess risk due to interaction; BMI, body mass index; LDL, Low-Density Lipoprotein cholesterol.

[a] Inverse probability of treatment weights was applied to adjust for baseline covariates (age, sex, country of birth, education levels, marital status, type of occupation, physical activity, smoking status, alcohol drinking, physical activity, BMI, waist circumference, hypertension, cardiovascular diseases, LDL cholesterol, statin prescription, elevated depressive symptoms at enrollment) and covariates at the first follow-up visit (BMI, waist circumference, hypertension, cardiovascular diseases, statin prescription). Inverse probability of censoring weights was also applied to adjust for right censoring at the first follow-up visit due to loss to follow-up.

[b] 1000 iterations were performed for bootstrapping to estimate 95% CI.

[c] The interaction was significant for cardiovascular mortality on both multiplicative and additive scales, and was significant for all-cause mortality on the additive scale. Multiplicative interaction was calculated by inserting multiplicative term between diabetes and elevated depressive symptoms ($HR_{DM(yes)\_Dep(yes)}$ /[$HR_{DM(yes)\_Dep(no)} \times HR_{DM(no)\_Dep(yes)}$]; null value = 1), and additive interaction was calculated by RERI ($HR_{DM(yes)\_Dep(yes)}$ - $HR_{DM(yes)\_Dep(no)}$ - $HR_{DM(no)\_Dep(yes)}$ +1; null value = 0).

Text S3.1. Creation of inverse-probability-treatment weights and inverse-probability-of-censoring weights for a marginal structural model under investigation

- Inverse-probability-of-treatment weights (IPTW) were calculated as follows:

$$\frac{pr[DM_1]}{pr[DM_1|\,Cov_1]} \times \frac{pr[Dep_2]}{pr[Dep_2|\,DM_1,Cov_1,Cov_2]}$$

where pr[.] is the probability function, DM_1 = diabetes at enrollment, Cov_1 = covariates at enrollment, Dep_2 = elevated depressive symptoms at the first follow-up visit, and Cov_2 = covariates at the first follow-up visit. Covariates at enrollment include age, sex, country of birth, education levels, marital status, type of occupation, physical activity, smoking status, alcohol drinking, physical activity, BMI, waist circumference, hypertension, cardiovascular diseases, LDL cholesterol, statin prescription, and elevated depressive symptoms. Covariates at the first follow-up visit (i.e., time-varying covariates) include BMI, waist circumference, hypertension, cardiovascular diseases, and statin prescription.

- Inverse-probability-of-censoring weights (IPCW) were calculated as follows:

$$\frac{pr[C_2=0]}{pr[C_2=0|\,DM_1,Cov_1]}$$

where C_2 = censoring (1, censored; 0, not censored) at the first follow-up visit, DM_1 = diabetes at enrollment, and Cov_1 = covariates at enrollment. Covariates at enrollment include age, sex, country of birth, education levels, marital status, type of occupation, physical activity, smoking status, alcohol drinking, physical activity, BMI, waist circumference, hypertension, cardiovascular diseases, LDL cholesterol, statin prescription, and elevated depressive symptoms.

- The final weights for each participant were created by multiplying the IPTW and the IPCW:

$$\frac{pr[DM_1]}{pr[DM_1|\,Cov_1]} \times \frac{pr[Dep_2]}{pr[Dep_2|\,DM_1,Cov_1,Cov_2]} \times \frac{pr[C_2=0]}{pr[C_2=0|\,DM_1,Cov_1]}$$

CHAPTER IV

Low HbA1c levels and all-cause or cardiovascular mortality among US adults without diabetes

4.1 Introduction

Glycated hemoglobin (HbA1c) is one of the major diagnostic biomarkers of diabetes, and it is well known that elevated HbA1c levels are associated with an increased risk of all-cause mortality as well as cardiovascular disease (CVD).[91–93] Additionally, previous studies have shown that relatively lower HbA1c levels are associated with increased risk of all-cause mortality and CVD among people with diabetes, suggesting a potential health burden for intensive treatment of glucose levels.[92–95] However, the effect of relatively lower HbA1c on long-term health outcomes among people without diabetes remains unclear as results from previous studies have been inconsistent.[93,96–102] Relatively lower HbA1c might be a proxy of malnutrition or an early stage of chronic disease.[93] Therefore, the observed increased risk of mortality may not reflect a causal effect of relatively lower HbA1c, but instead, be a reflection of the underlying poor health. Moreover, although the risk of mortality according to HbA1c levels may vary over time, previous studies have employed a Cox proportional hazard model to estimate hazard ratios, potentially violating *proportional hazards assumption* (i.e. relative hazards are assumed to not vary over time). In this context, to investigate the impact of relatively lower HbA1c on long-term health outcomes, analyses using flexible models that consider an array of confounders not previously accounted for and that account for time-varying risk in the estimation, are needed.

One of the major impediments for effectively addressing the causal pathways from HbA1c to mortality is the complex multifactorial interactions between blood glucose levels and sociological, biological, and clinical factors. Ample evidence exists that numerous factors are associated with both HbA1c and mortality, including demographics, socioeconomic status, diet,

exercise, biomarkers, comorbidities, and medication.[10,103] Due to such a high-dimensional data structure, it has often been challenging to integrate all this information and accurately establish a causal relationship between relatively lower HbA1c and adverse health outcomes. Furthermore, given that interventions using a clinical trial approach to lower glucose levels among people without diabetes would not be ethical and feasible, causal analyses using observational data are needed on this topic.

In recent years, there have been substantial advances in the application of machine learning algorithms within the framework of causal inference including the g-formula,[104] propensity scores,[105] and targeted maximum likelihood estimation.[106] For example, the g-formula framework allows the researcher to build an outcome prediction model based on observed quantities, and then predict potential outcomes under hypothetical exposure levels.[107] Given the rapidly expanding availability of data, flexible machine learning algorithms may offer advantages in applying this step of the g-formula to efficiently specify the prediction model, as they have the ability to discover whether it is important to include interactions, non-linear, and higher-order effects which may not be easily covered by conventional regression models.[107,108]

In this study, using machine learning algorithms as well as conventional logistic regression within the parametric g-formula, we estimated the effect of relatively lower HbA1c on all-cause and cardiovascular mortality among US general adults.

4.2 Methods

4.2.1 Study Design and Participants

We used data from the U.S. National Health and Nutrition Examination Survey (NHANES) 1999-2014.[109] NHANES is a large-scale, multistage, nationally representative survey of the civilian noninstitutionalized population in the United States, conducted by the National Center for Health Statistics (NCHS). Structured interview data and physical examination results, including urine and/or blood samples, are collected continuously and released in two-year cycles.[109] All participants provided informed written consent at enrollment and completed a household interview followed by a physical examination in a mobile examination center. The response rates of NHANES during the study period were 70-80%.[110] The study protocols of NHANES were approved by the NCHS Institutional Review Board.[111]

There were 39,520 participants aged ≥20 years at enrollment for whom HbA1c was available. We excluded participants with extremely low HbA1c levels (<4.0%) that could be induced by severe liver disease or hemolysis (n=18).[112] We further excluded participants who lacked time-to-event data for death due to insufficient identifying information when linking the mortality data (n=49). The final analytical cohort contained 39,453 participants.

4.2.2 Measurement of Variables

4.2.2.1 Exposure and diagnosed diabetes ascertainment

During visits to the mobile examination center, phlebotomists obtained blood samples from participants according to a standardized protocol after participants fasted at least 8 hours and no more than 24 hours. These samples were subsequently analyzed to measure HbA1c using high-performance liquid chromatography.[113] We stratified participants with HbA1c within the normal range into three groups by HbA1c levels as follows: low HbA1c, 4.0 to <5.0%; mid-level HbA1c

(referent group), 5.0 to <5.7%; and prediabetes, 5.7 to <6.5% as done in previous studies.[96,114] We also categorized participants with HbA1c ≥6.5% or taking antihyperglycemic therapies and insulin into the "diabetes" group. We included them in our analysis as a positive control group for whom an increased risk of mortality is expected compared with the mid-level HbA1c group.

4.2.2.2 Covariates

Demographic variables included age, sex (male, female), race/ethnicity (non-Hispanic White, non-Hispanic Black, Mexican-American or others), citizenship status (US or other), educational status (less than high school, high school or General Education Degree, or more than high school), health insurance status (private, public, none), marital status (single, married), and the poverty–income ratio (the ratio of the family income to the poverty threshold; range, 0-5). Smoking status (never, current, former) and physical activity levels (≥moderate or not) were self-reported. Diet information was obtained from 24-hour dietary recall collected by trained interviewers using a computer-based interactive platform (Table S4.1). As comorbidities, we selected anemia, angina, arthritis, asthma, cancer, chronic heart failure, emphysema, heart attack, hypertension, liver failure, and stroke (self-reported). The use of statins, antihypertensives, and antidepressants was also self-reported. Biomarkers were measured according to NHANES laboratory procedure manuals (Table S4.1).[115] All covariates were measured or reported at enrollment.

4.2.2.3 Outcomes ascertainment

We used the NCHS Public-Use Linked Mortality File through December 31, 2015, to ascertained death certificate information provided by the National Death Index (NDI)[116] through record

matching by social security number, name, date of birth, race/ethnicity, sex, state of birth, and state of residence. The primary outcome for the present study was all-cause mortality, and the secondary outcome was cardiovascular mortality. The cause of death was determined based on the International Classification of Diseases, Tenth version (ICD–10). Cardiovascular disease was classified using ICD–10 codes I00–09, I11, I13, I20–51, and I60–69.[117]

4.2.3 Statistical analyses

We employed the parametric g-formula algorithm, a generalization of the method known as standardization, to estimate the risk of death at 5 and 10 years for each HbA1c category. All models included continuous and quadratic terms for the follow-up year since NHANES enrollment, HbA1c category, an indicator variable for NHANES enrollment year, and all of the above-mentioned covariates. Missing data among covariates (28% of all participants had at least one missing value) were imputed with a random forest approach.[118]

In the parametric g-formula, we first fitted outcome prediction models using the exposure (HbA1c categories) and the above-mentioned 72 covariates after arranging data into a person-time structure. To find the best predictive model for the outcome in this first step of the parametric g-formula, we developed a reference model and three machine learning models for each outcome using a training set (composed of a randomly selected 50% of the data). As the reference model, we fitted a pooled logistic regression model. In this model, we pooled observations from each follow-up year into a single dataset and employed logistic regression to predict the occurrence of each outcome. We also fitted tree-based machine learning algorithms (random forest[119] and gradient-boosted decision tree[120]) and SuperLearner.[121] To minimize the

potential for overfitting, we performed 10-fold cross-validation for each model. After developing these prediction models, we computed the following prediction performance measures for each model in the test set (50% randomly selected samples): the area under the receiver-operating-characteristics curve (AUC) and confusion matrix results (i.e., sensitivity, specificity, positive predictive value, and negative predictive value).

As the second step of the parametric g-formula, using the total sample, we employed the pooled logistic regression model and one of the machine learning algorithms with the best prediction performance and predicted the values for the potential outcomes under counterfactual exposures. Then, we estimated the average marginal effect of exposure on the outcome. We compared the estimated risk of death at 5 and 10 years had all eligible participants belonged to each of the HbA1c categories using a risk ratio (RR) and a risk difference (RD) measure.[122] Robust 95% confidence intervals (CIs) were estimated by repeating these analyses on 200 bootstrapped samples. A more detailed discussion and coding for the parametric g-formula are presented elsewhere.[123] To evaluate mortality risks according to continuous HbA1c, we also employed restricted cubic spline models fitted with Cox proportional hazard regression with 3 knots (10^{th}, 50^{th}, and 90^{th} percentile).

The stratum-specific analyses were conducted by age: younger (<65 years) versus older (≥65 years) and by sex: male versus female. The P-value for heterogeneity was calculated using the method proposed by Altman and Bland.[124] We also performed the following sensitivity analyses to assess the robustness of our findings: 1) we performed complete case analysis with NHANES survey weights to account for unequal probabilities of selecting NHANES participants and

nonresponse of those eligible and approached (n=28,312),[125] and 2) we re-analyzed data restricting participants to those with hemoglobin ≥13 g/dl in males and hemoglobin ≥12 g/dl in females who did not report anemia because *HbA1c* could be affected by *anemia* (n=34,740).[126,127] All statistical analyses were conducted using R version 3.5.2.

4.3 Results

4.3.1 Demographic characteristics

The mean age of participants was 49.5 years (standard deviation, 18.3; median, 48; interquartile range, 34 to 64), and 48.1% were male. Demographic characteristics of participants across HbA1c groups are shown in Table 4.1 and Table S4.1.

4.3.2 Prediction of all-cause mortality and cardiovascular mortality

Overall, the median duration of follow-up was 7.5 years, and 5,118 all-cause deaths and 1,116 cardiovascular deaths were identified (Table S4.2). Kaplan-Meier survival curves by HbA1c levels are shown in Figure S4.1. All of the candidate algorithms, including the pooled logistic regression model, showed high prediction performance of all-cause mortality and cardiovascular mortality with AUCs ranging from 0.86 to 0.90 (Table 4.2). Sensitivity, specificity, and negative predictive value were also similar for pooled logistic regression and SuperLearner, while random forest yielded a lower specificity and gradient boosting yielded a lower sensitivity.

4.3.3 Estimated risk of all-cause mortality and cardiovascular mortality at 5 and 10 years

After adjusting for all potential confounders including demographic characteristics, diet, exercise, comorbidities, biomarkers, and medications using a pooled logistic regression model

within the parametric g-formula, compared to the mid-level HbA1c group, the low HbA1c group showed a 30% (95% CI, 16 to 48) and a 12% (95% CI, 3 to 22) increased risk of all-cause mortality at 5 years and 10 years of follow-up, respectively. On the absolute scale, the low HbA1c group showed a 1.83 (95% CI, 1.02 to 2.97) and a 1.66 (95% CI, 0.35 to 3.00) percentage points increase for all-cause mortality risk at 5 and 10 years of follow-up, respectively (Figure 4.1, Table 4.3). We found no evidence of an association between HbA1c and cardiovascular mortality. The findings were qualitatively consistent when we used SuperLearner (which showed the highest predictive performance among the three machine learning algorithms) within the parametric g-formula (Figure 4.1).

We did not find evidence for an association in the prediabetes group with all-cause mortality at 5 and 10 years of follow-up (Figure 4.1, Table S4.3). As expected, the diabetes group showed an increased risk of all-cause and cardiovascular mortality regardless of model specifications (i.e., either the pooled logistic regression model or SuperLearner) (Figure 4.1, Table S4.4). These associations were also found when restricted cubic spline curves were fitted with Cox proportional hazard regression (Figure S4.2).

4.3.4 Stratum specific analysis by age and sex

We found similar associations for low HbA1c and all-cause mortality in the younger and the older population on the relative risk scale (Figure 4.2, Table S4.5). When we stratified by sex, we estimated increased risk of all-cause mortality for low HbA1c among females but not among males (Figure 4.2, Table S4.5). We found no evidence for an association between low HbA1c and cardiovascular mortality in any subgroups stratified by age and sex (Table S4.6).

4.3.5 Sensitivity analyses

The results for all-cause mortality did not substantially change when we performed complete-case analysis using NHANES survey weights (Table S4.7) and when we re-analyzed the data after restricting to participants without anemia, particularly at 5 years (Table S4.8).

4.4 Discussion

Using a nationally representative database of US adults and controlling for potential confounders (e.g., demographic characteristics, diet, exercise, comorbidities, biomarkers, and medications) with several statistical algorithms, we found that individuals with low HbA1c (4.0 to <5.0%) were experiencing an increased risk of all-cause mortality at 5- and 10- years of follow-up compared to those with mid-level HbA1c (5.0 to <5.7%). This relationship was stronger for females than males. We found no evidence that low HbA1c was associated with cardiovascular mortality.

The question of whether relatively lower HbA1c is beneficial or harmful for people without diabetes has been actively debated for a long time. Although some previous studies have reported associations between relatively lower HbA1c and increased CVD and mortality,[97–100] its clinical and biological relevance has remained unclear. As HbA1c is becoming more frequently (and routinely) measured based on clinical guidelines,[128] the chance that clinicians detect people with relatively lower HbA1c might increase, and there is a need to answer this long-debated question about the potential burden of relatively lower HbA1c levels on health. In this context, our findings provide new evidence, indicating that we may need to carefully monitor people with relatively lower HbA1c.

Our findings were consistent with previous cohort studies investigating the association between low HbA1c and all-cause mortality among people without diabetes.[97,99,100] A recent study among U.S. adults aged ≥50 years without diabetes from the Health and Retirement Study also reported a reverse J-shaped association between HbA1c and all-cause mortality, but not cardiovascular mortality.[93] Given the null association between relatively lower HbA1c and all-cause mortality over 15 years of follow-up among Japanese adults,[101] the association might be heterogeneous across race/ethnicity, which requires further investigation. A previous study in NHANES also did not find an association between relatively lower HbA1c and all-cause mortality over a median follow-up of 9 years, but this analysis only included 7,333 participants aged ≥65 years enrolled before 2004.[96] Furthermore, most of these studies have suffered from limitations including a limited number of covariates adjusted for (i.e., unmeasured confounding bias), violations of proportional hazard assumptions, or a relatively short follow-up period. Our study, using high-dimensional data from a national survey with long follow-up time and flexible statistical modeling, overcomes some of these limitations, and therefore, helps to advance the current state of knowledge about the potential impact of low HbA1c on mortality.

Underlying biologic mechanisms for the association between relatively lower HbA1c and mortality have still not been established. Poor health status (e.g., malnutrition, unfavorable profiles of red blood cell-related factors, inflammation, decreased liver function, or an early stage of chronic disease) has been proposed as an explanation for the association between relatively lower HbA1c and mortality among people without diabetes.[98,112,114] Hypoglycemia induces sympathoadrenal activation, inflammation, and endothelial dysfunction, all of which could lead

to chronic and cardiometabolic diseases.[129] Given these proposed mechanisms and the fact that we did not find evidence for an association between low HbA1c and cardiovascular mortality (likely due to insufficient statistical power), further investigations with a larger sample size focusing on a high CVD risk population are warranted. Furthermore, we found a stronger association among females than males, particularly at 5 years follow-up. This was mainly due to the higher mortality risk for mid-level HbA1c among males than females, although both sexes showed similar mortality risks at low HbA1c. Given sex differences in glucose metabolism,[130] our findings also indicate the importance of evaluating HbA1c by sex.

As expected based on ample prior evidence,[131–133] the diabetes group showed an increased risk of all-cause mortality. A cohort study of one million US adults reported diabetes to be associated with a higher risk of mortality for several diseases such as CVD, cancer, respiratory dysfunction, digestive diseases, genitourinary disorders, and even accidents.[131] Although mortality and incidence of cardiovascular events among people with diabetes have decreased over the last two decades, mainly owing to remarkable advancements in the treatment of CVD and diabetes,[132,133] our findings highlight that there is still a need for further improvement of diabetes management to avoid complications.

The present study has three major strengths. First, our study utilized a large, nationally representative sample of the US general population with linkage to the most updated national mortality database. Second, we applied the parametric g-formula that does not require the proportional hazard assumption and allows us to estimate clinically meaningful absolute/relative risks.[123] This approach also enabled us to estimate risks at different time points (i.e., 5- and 10-

years of follow-up). Lastly, we employed several ensemble machine learning algorithms to build the outcome prediction models in the first step of the parametric g-formula, including an ensemble method called SuperLearner that combines multiple machine learning algorithms with weights estimated to maximize performance.[121] Our results suggest that a conventional logistic regression modeling approach, which has a much lower computational cost than machine learning algorithms, may well suffice to answer our research question as risk factors are well-known. But even in a scenario such as ours, comparing the findings from both logistic regression models and machine learning algorithms provides a transparent approach to addressing the potential for bias due to model misspecification.

Our study has limitations. First, although we included an extensive set of covariates, there is always a possibility for bias due to unmeasured confounding in observational study design. For example, low-density lipoprotein cholesterol levels were not available for many NHANES participants; therefore, we used statin use as a proxy for dyslipidemia. The lack of detailed information on diabetes such as family history and antibodies may also limit the interpretation of diabetes and mortality association, but does not affect the interpretation of our primary outcomes (i.e., low HbA1c and mortality among people without diabetes). The wide range of age in the present study may also raise a concern about residual confounding by age even after adjusting for age. However, as increasing age is negatively associated with low HbA1c and positively associated with mortality risks, such bias would be expected to cause an underestimation of the effect of low HbA1c on mortality. Given that clinical trials may not be feasible and ethical on this topic, future studies using other epidemiological approaches such as Mendelian randomization should be considered to validate our findings. Second, diet, lifestyle, and

comorbidities may have been mismeasured because these variables were self-reported. Third, as HbA1c was only measured at baseline, we had no information about how changes in HbA1c may or may not contribute to the increased risk of mortality. Lastly, covariate information was also only available at baseline. Thus, the exposure-confounders relationships were not well defined temporarily, and we cannot rule out the possibility of reverse causation and over-adjustment. Further longitudinal studies with measurements of HbA1c and other covariates at multiple time points are needed to overcome this limitation.

4.5 Conclusion

Low HbA1c was associated with an increased risk of all-cause mortality at 5 and 10 years of follow-up among US adults. Our findings may indicate the importance of carefully monitoring individuals having relatively lower HbA1c without diabetes as well as individuals with diabetes in clinical practice. A better understanding of this relationship would enable healthcare professionals to design effective public health interventions to reduce the risk of long-term adverse health outcomes that may be related to relatively lower HbA1c.

4.6 Tables and Figures

Table 4.1 Baseline characteristics according to glycated hemoglobin (HbA1c) levels in NHANES 1999-2014 [a, b]

	HbA1c levels within the normal range			HbA1c≥6.5% or antidiabetic medication
	Low (4.0 to <5.0%)	Mid-level (5.0 to <5.7%)	Prediabetes (5.7 to <6.5%)	
Total, n	4,314	23,953	5,921	5,265
HbA1c %, median (IQR)	4.8 (4.7-4.9)	5.4 (5.2-5.5)	5.9 (5.8-6.1)	7.0 (6.5-8.2)
Age (years)	36.5 ± 15.0	46.8 ± 17.9	59.1 ± 15.6	61.7 ± 13.4
Sex (female), %	59.1	51.7	50.4	48.5
Race/ethnicity, %				
Non-Hispanic White	54.0	50.1	40.6	36.3
Non-Hispanic Black	17.4	17.2	27.0	26.8
Mexican-American	16.3	18.3	16.4	21.4
Others	12.3	14.4	16.0	15.5
Education status, %				
Less than 9^{th} grade	7.3	10.7	16.4	22.0
9^{th}-11^{th} grade	13.7	15.1	17.1	19.3
High school or GED	21.4	23.4	24.5	22.3
Higher than high school	57.6	50.8	42.0	36.4
Married, %	49.7	53.6	54.5	56.0
Smoking, %				
Never	59.4	54.2	49.8	48.8
Current	21.9	22.6	20.6	16.7
Former	18.7	23.2	29.6	34.5
Insurance status, %				
Private	33.1	34.0	38.5	32.9
Public	42.1	42.1	41.5	52.3
Uninsured	24.8	23.9	20.0	14.8
Poverty-income ratio	2.64 ± 1.66	2.62 ± 1.65	2.45 ± 1.57	2.23 ± 1.50
Physical activity levels (≥moderate), %	68.7	65.2	56.8	47.6
Anemia, %	4.7	3.3	3.5	6.3
Angina, %	0.7	2.1	3.9	7.8
Arthritis, %	12.2	22.5	36.4	44.3
Asthma, %	14.0	12.6	12.3	14.6
Cancer, %	4.2	8.1	12.2	12.9
Chronic Heart failure, %	1.0	1.9	4.1	9.8
Emphysema, %	0.6	1.6	3.1	3.6
Heart attack, %	1.3	2.9	6.2	11.1
Hypertension, %	15.9	27.2	47.1	65.5
Liver failure, %	2.9	3.1	3.6	5.8
Stroke, %	1.7	2.6	5.2	8.7
Statin use, %	2.8	9.7	23.8	42.3
Antihypertensive use	8.2	18.1	38.3	58.2
Antidepressant use, %	7.9	8.8	9.6	14.5

Systolic blood pressure (mmHg)	116.8 ± 16.5	123.3 ± 19.3	131.9 ± 20.8	134.2 ± 21.4
Diastolic blood pressure (mmHg)	67.8 ± 12.9	70.8 ± 12.9	71.7 ± 14.1	68.8 ± 15.6
Waist (cm)	91.6 ± 14.1	96.1 ± 14.9	103.4 ± 15.2	108.9 ± 15.6
BMI (kg/m^2)	26.5 ± 5.5	28.0 ± 6.2	30.5 ± 6.9	32.4 ± 7.4

Abbreviations: NHANES, National Health and Nutrition Examination Survey; IQR, interquartile range; GED, General Educational Development; BMI, body mass index.

[a] Data are presented as count (percentage) or mean ± SD, otherwise specified.

[b] Other variables (dietary information and biomarkers) are shown in Supplementary Table S4.1.

Table 4.2 Predictive ability of pooled logistic regression model, tree-based algorithms, and SuperLearner for all-cause and cardiovascular mortality.

Models	**AUC**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Outcome: All-cause mortality					
Logistic regression model	0.87	0.80	0.78	0.05	>0.99
Random forest	0.86	0.91	0.54	0.03	>0.99
Gradient Boosting	0.86	0.54	0.92	0.09	>0.99
SuperLearner	0.87	0.85	0.75	0.04	>0.99
Outcome: Cardiovascular mortality					
Logistic regression model	0.90	0.84	0.82	0.01	>0.99
Random forest	0.88	0.94	0.61	0.01	>0.99
Gradient Boosting	0.89	0.42	0.96	0.03	>0.99
SuperLearner	0.90	0.90	0.74	0.01	>0.99

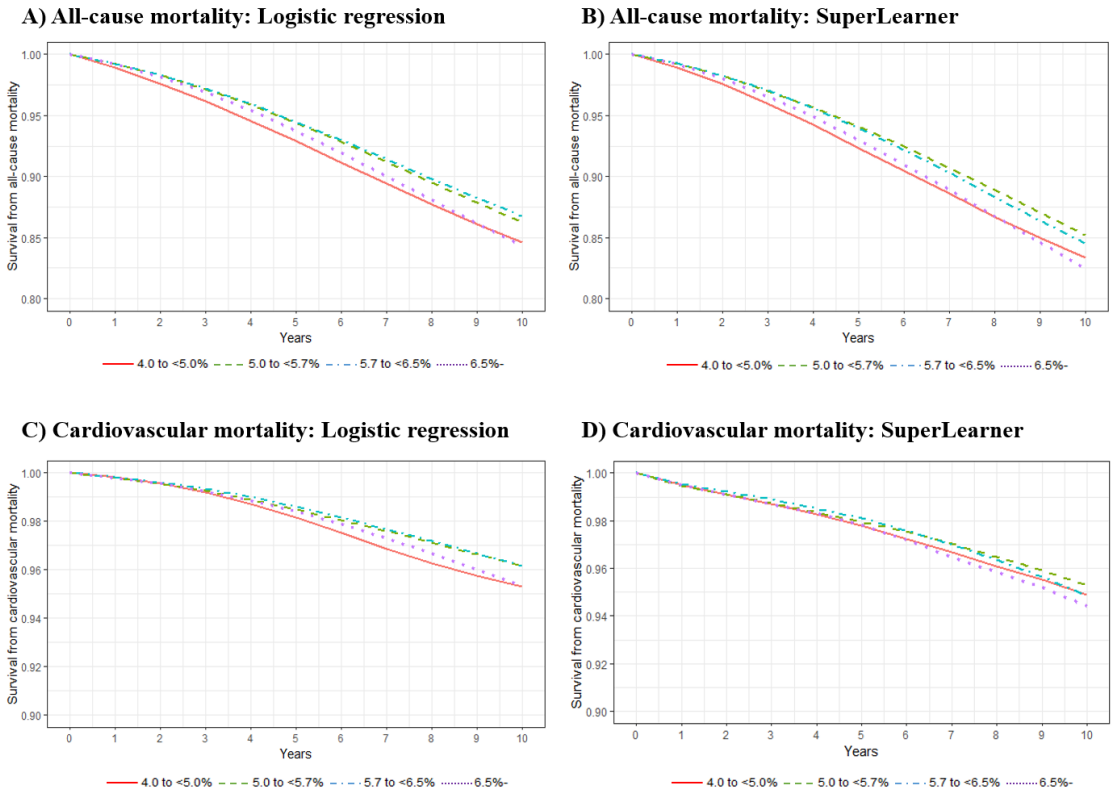
AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value
Each model included all 72 covariates listed in Table 4.1 and Table S4.1 (e.g., demographic characteristics, diet, exercise, comorbidities, biomarkers, and medications). Confusion matrix results (i.e. sensitivity, specificity, PPV, and NPV) were at the cutoff value of the prevalence of each outcome. PPVs were generally low for all algorithms due to the small number of outcomes overall (i.e., all-cause and cardiovascular mortality).

Table 4.3. Adjusted all-cause and cardiovascular mortality risk ratio and risk difference at 5 and 10 years among participants with low glycated hemoglobin (HbA1c; 4.0 to <5.0%) compared to those with mid-level HbA1c (5.0 to <5.7%) using parametric g-formula with pooled logistic regression models. [a]

Outcomes	**Follow-up periods**	
	5 years	**10 years**
All-cause mortality		
Number of events/Total number of participants among low HbA1c group	149/3453 (4.3%)	249/2205 (11.3%)
Number of events/Total number of participants among mid-level HbA1c group	1125/18024 (6.2%)	2104/10906 (19.3%)
Adjusted risk ratio (95% CI)	1.30 (1.16 to 1.48)	1.12 (1.03 to 1.22)
Adjusted risk difference (95% CI)	+1.83% (1.02 to 2.97)	+1.66% (0.35 to 3.00)
Cardiovascular mortality		
Number of events/Total number of participants among low HbA1c group	23/3327 (0.7%)	46/2002 (2.3%)
Number of events/Total number of participants among mid-level HbA1c group	263/17162 (1.5%)	452/9254 (4.9%)
Adjusted risk ratio (95% CI)	1.17 (0.80 to 1.59)	1.21 (0.92 to 1.54)
Adjusted risk difference (95% CI)	+0.28% (-0.35 to 0.99)	+0.83% (-0.32 to 2.01)

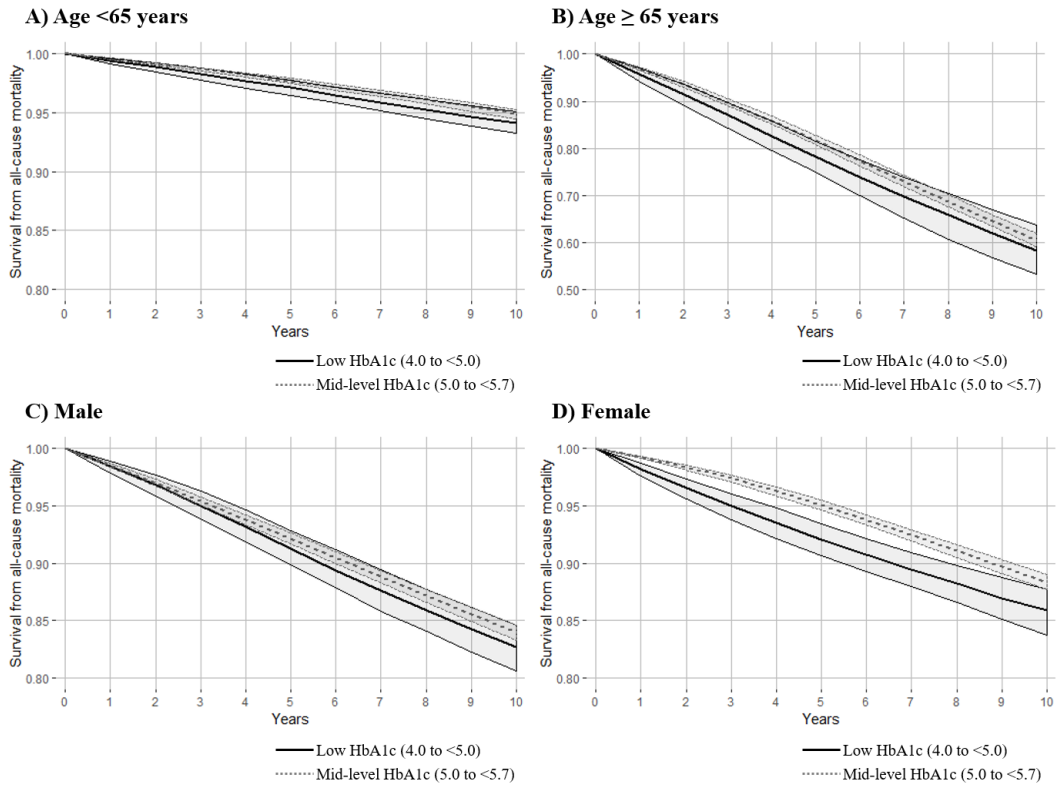
[a]200 iterations were performed for bootstrapping to estimate 95% confidence interval.

Figure 4.1 Adjusted all-cause and cardiovascular mortality risk according to glycated hemoglobin (HbA1c) levels using parametric g-formula with pooled logistic regression models and SuperLearner.



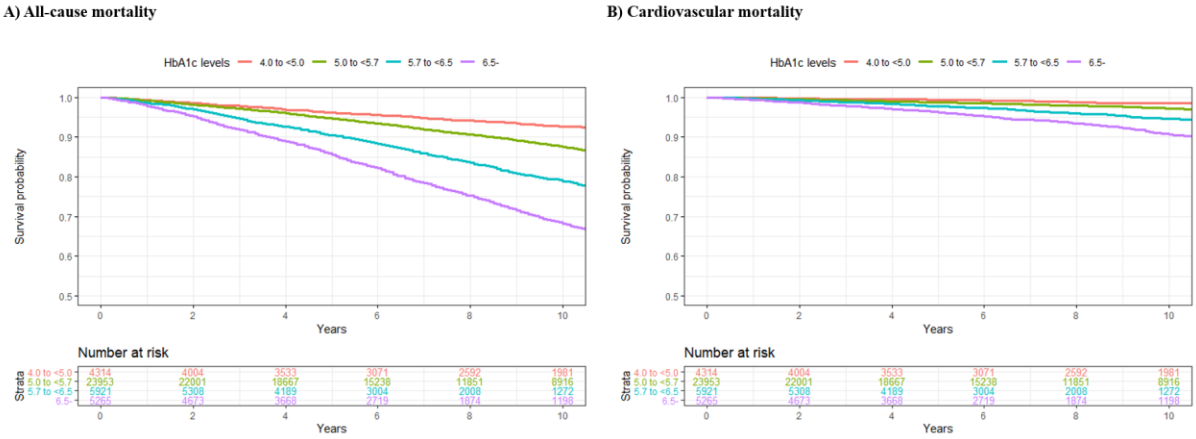
The ranges of the survival rate (Y-axis) presented in figures were 0.8 to 1.0 for all-cause mortality and 0.9 to 1.0 for cardiovascular mortality. Robust 95% confidence intervals (CIs) for each HbA1c category estimated by bootstrapping (in the pooled logistic regression model**)** are presented in Table 4.3, and Table S4.3 and S4.4.

Figure 4.2 Adjusted all-cause mortality risk among participants low glycated hemoglobin (HbA1c; 4.0 to <5.0%) and mid-level HbA1c (5.0 to <5.7%) stratified by age and sex using parametric g-formula with pooled logistic regression models.



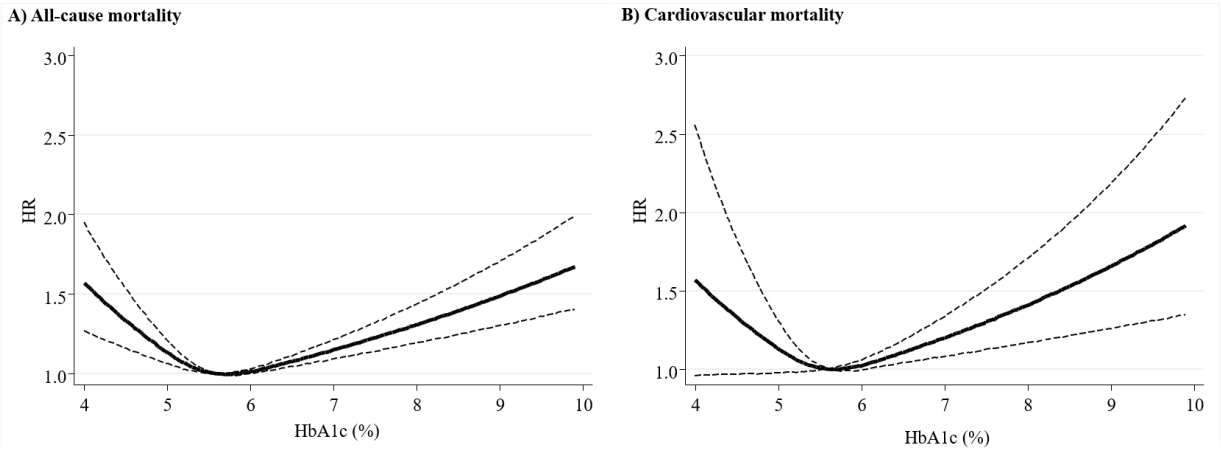
The ranges of the survival rate (Y-axis) presented in figures were 0.8 to 1.0 except for the older population (range 0.5 to 1.0) who had a higher mortality rate than other groups. Robust 95% confidence intervals (CIs) were estimated by repeating these analyses on 200 bootstrapped samples.

Figure S4.1 Kaplan-Meier survival curve for all-cause and cardiovascular mortality



Kaplan-Meier survival curve shows the unadjusted survival probability in each HbA1c category.

Figure S4.2 Association of HbA1c with all-cause and cardiovascular mortality using restricted cubic spline curve fitted with Cox proportional hazard regression.



Y-axis shows HR of (A) all-cause and (B) cardiovascular mortality adjusted for 72 variables including demographic characteristics, diet, exercise, biomarkers, and comorbidities. The dashed lines represent the confidence intervals for the restricted cubic spline model (reference is 5.6%). The range of HbA1c was restricted to ≤10% because of too few data points >10%.

Table S4.1 Baseline characteristics of all covariates not described in Table 1 (diet information and biomarkers: alphabetical order) according to HbA1c levels in NHANES 1999-2014 [a]

	HbA1c levels within the normal range			HbA1c≥6.5% or antidiabetic medication
	Low (4.0 to <5.0%)	Mid-level (5.0 to <5.7%)	Prediabetes (5.7 to <6.5%)	
Total, n	4,314	23,953	5,921	5,265
Diet information [b]				
Alcohol (gm)	13.8 ± 35.2	11.2 ± 31.0	7.0 ± 23.0	4.9 ± 20.9
Caffeine (mg)	132.1 ± 185.6	164.6 ± 222.8	152.1 ± 195.8	144.5 ± 197.3
Calcium (mg)	946.8 ± 619.4	909.5 ± 611.9	825.8 ± 520	803.6 ± 511.6
Carbohydrate(gm)	275.9 ± 132.8	267.5 ± 133.8	245.2 ± 117.1	219.2 ± 110.1
Dietary fiber (gm)	16.3 ± 10.8	16.5 ± 10.4	15.8 ± 9.5	16.0 ± 10.2
Energy (kcal)	2240.9 ± 1041.2	2177.4 ± 1039.6	1977.7 ± 902	1828.9 ± 875.5
Iron (mg)	15.7 ± 9.2	15.3 ± 9.1	14.2 ± 8.1	14.2 ± 8.2
Magnesium (mg)	296.9 ± 162.3	293.7 ± 151.7	272.8 ± 133.0	266.6 ± 137.8
Phosphorus (mg)	1371.7 ± 708.0	1348 ± 692.4	1235.3 ± 605.4	1212.8 ± 612.4
Potassium (mg)	2694.0 ± 1392.9	2699.1 ± 1320.2	2546.0 ± 1190.3	2483.8 ± 1183.0
Protein (gm)	83.6 ± 43.9	82.4 ± 43.9	76.1 ± 39.3	75 ± 38.7
Total fat (gm)	81.7 ± 46.9	80.9 ± 47.4	74.4 ± 42.8	71.4 ± 42.9
Vitamin B1 (Thiamin) (mg)	1.7 ± 1.0	1.6 ± 1.0	1.5 ± 0.8	1.5 ± 0.8
Vitamin B2 (Riboflavin) (mg)	2.1 ± 1.2	2.1 ± 1.3	1.9 ± 1.2	1.9 ± 1.1
Vitamin B3 (Niacin) (mg)	25.2 ± 14.7	24.7 ± 15.2	22.7 ± 13.6	21.8 ± 12.3
Vitamin B6 (mg)	2.1 ± 1.4	2.0 ± 1.5	1.9 ± 1.3	1.8 ± 1.2
Vitamin B9 (Folate) (mcg)	414.6 ± 256.0	403.2 ± 253.2	375.1 ± 226.4	364.8 ± 222.8
Vitamin B12 (mcg)	5.3 ± 6.9	5.2 ± 7.1	4.9 ± 9.8	4.8 ± 7.3
Vitamin C (mg)	100.4 ± 114.9	91.4 ± 104.5	86.1 ± 92.8	82.2 ± 95.2
Zinc (mg)	12.0 ± 8.4	11.8 ± 9.7	10.7 ± 7.5	10.6 ± 8.9
Biomarkers				
Alanine aminotransferase ALT (IU/L)	24.0 ± 27.9	25.1 ± 22.5	26.4 ± 31.4	27.1 ± 29.3
Albumin (g/dL)	4.3 ± 0.5	4.3 ± 0.4	4.2 ± 0.3	4.1 ± 0.3
Alkaline phosphatase (IU/L)	68.6 ± 29.6	70 ± 26.8	73.9 ± 25.2	77.7 ± 32.0
Aspartate aminotransferase AST (IU/L)	25.3 ± 22.2	25.4 ± 17.2	26.5 ± 25.3	26.4 ± 18.3
Bicarbonate (mmol/L)	24.1 ± 2.5	24.7 ± 2.3	25.1 ± 2.3	24.9 ± 2.5
Blood urea nitrogen (mg/dL)	11.3 ± 5.0	12.9 ± 5.2	14.4 ± 6.3	16.5 ± 8.9
Chloride (mmol/L)	103.9 ± 2.8	103.7 ± 2.8	103.6 ± 3.0	102.5 ± 3.5
Cholesterol (mg/dL)	189.6 ± 42.6	198.6 ± 41.0	202.7 ± 43.2	190.2 ± 48.1
Creatinine (mg/dL)	0.8 ± 0.5	0.9 ± 0.4	0.9 ± 0.4	1.0 ± 0.7
Direct HDL-Cholesterol (mg/dL)	56.7 ± 17.5	54 ± 16.3	50.4 ± 14.6	47.4 ± 13.6
Gamma glutamyl transferase (U/L)	28.9 ± 59.5	27.9 ± 42.0	31.8 ± 34.5	38.3 ± 55.9
Lactate dehydrogenase (U/L)	126.8 ± 34.7	131.6 ± 31.5	139 ± 33.5	137.6 ± 36.2
Phosphorus (mg/dL)	3.7 ± 0.6	3.7 ± 0.6	3.7 ± 0.6	3.7 ± 0.6
Potassium (mmol/L)	3.9 ± 0.3	4.0 ± 0.3	4.0 ± 0.4	4.1 ± 0.4
Sodium (mmol/L)	139.0 ± 2.4	139.2 ± 2.3	139.4 ± 2.4	138.7 ± 2.8
Total bilirubin (mg/dL)	0.8 ± 0.4	0.7 ± 0.3	0.7 ± 0.2	0.7 ± 0.3
Total calcium (mg/dL)	9.4 ± 0.4	9.4 ± 0.4	9.4 ± 0.4	9.5 ± 0.4
Total protein (g/dL)	7.2 ± 0.5	7.2 ± 0.5	7.2 ± 0.5	7.2 ± 0.5

Triglycerides (mg/dL)	127.7 ± 105.9	141.9 ± 117.7	166.9 ± 123.8	204.3 ± 206.0
Uric acid (mg/dL)	5.0 ± 1.5	5.3 ± 1.4	5.8 ± 1.4	5.7 ± 1.6
White blood cell count (1000 cells/uL)	7.2 ± 2.5	7.2 ± 2.3	7.4 ± 2.9	7.6 ± 2.3
Hemoglobin (g/dL)	14.0 ± 1.7	14.2 ± 1.5	14.1 ± 1.5	13.8 ± 1.6
Hematocrit (%)	41.0 ± 4.7	41.9 ± 4.4	41.6 ± 4.2	40.8 ± 4.6
Albumin, urine (ug/mL)	25.2 ± 175.3	26.6 ± 186.5	36.0 ± 170.0	184 ± 887.0
Creatinine, urine (mg/dL)	132.4 ± 87.1	127.2 ± 81.5	122.8 ± 76.3	115.9 ± 73.0
NHANES cycle [c]				
1999-2002	30.3	24.0	15.5	18.8
2003-2006	27.6	24.0	16.9	20.2
2007-2010	21.9	26.7	35.2	31.6
2011-2014	20.3	25.2	32.4	29.5

Abbreviations: NHANES, National Health and Nutrition Examination Survey; GED, General Educational Development; BMI, body mass index.

[a] Data are presented as count (percentage) or mean ± SD

[b] The higher intake in lower HbA1c groups than higher HbA1c groups was generally reflected by the difference in age and other demographic distribution as shown in Table 1.

[c] NHANES HbA1c data were standardized by participating in the National Glycohemoglobin Standardization Program (NGSP).

Table S4.2 The underlying cause of death across the study population, NHANES 1999-2014 linked to mortality data through 2015.

The underlying cause of death	**N (%)**
Cardiovascular disease including stroke	1,116 (21.8%)
Cancer	1,078 (21.0%)
Chronic lower respiratory diseases	199 (3.9%)
Accidents	147 (2.9%)
Diabetes mellitus	132 (2.6%)
Alzheimer's disease	126 (2.5%)
Nephritis, nephrotic syndrome	89 (1.7%)
Influenza and pneumonia	84 (1.6%)
Others/unknown	2,147 (42.0%)
Total	**5,118 (100%)**

Table S4.3 Adjusted all-cause and cardiovascular mortality risk ratio and risk difference at 5 and 10 years among participants with prediabetes (5.7 to <6.5%) compared to those with mid-level HbA1c (5.0 to <5.7%) at 5 and 10 years using parametric g-formula with pooled logistic regression models. [a]

Outcomes	**Follow-up periods**	
	5 years	**10 years**
All-cause mortality		
Number of events/Total number of participants among prediabetes group	483/4077 (11.9%)	787/2030 (38.8%)
Number of events/Total number of participants among mid-level HbA1c group	1125/18024 (6.2%)	2104/10906 (19.3%)
Adjusted risk ratio (95% CI)	0.97 (0.91 to 1.05)	0.95 (0.90 to 1.01)
Adjusted risk difference (95% CI)	-0.17% (-0.62 to 0.31)	-0.69% (-1.46 to 0.12)
Cardiovascular mortality		
Number of events/Total number of participants among prediabetes group	109/3703 (2.9%)	180/1423 (12.7%)
Number of events/Total number of participants among mid-level HbA1c group	263/17162 (1.5%)	452/9254 (4.9%)
Adjusted risk ratio (95% CI)	0.92 (0.78 to 1.11)	0.99 (0.83 to 1.14)
Adjusted risk difference (95% CI)	-0.14% (-0.39 to 0.17)	-0.03% (-0.64 to 0.53)

[a]200 iterations were performed for bootstrapping to estimate 95% confidence interval.

Table S4.4 Adjusted all-cause and cardiovascular mortality risk ratio and risk difference at 5 and 10 years among participants with diabetes (HbA1c≥6.5%) compared to those with mid-level HbA1c (5.0 to <5.7%) using parametric g-formula with pooled logistic regression models. [a]

Outcomes	**Follow-up periods**	
	5 years	**10 years**
All-cause mortality		
Number of events/Total number of participants among diabetes group	662/3828 (17.3%)	1126/2302 (48.9%)
Number of events/Total number of participants among mid-level HbA1c group	1125/18024 (6.2%)	2104/10906 (19.3%)
Adjusted risk ratio (95% CI)	1.10 (1.03 to 1.19)	1.13 (1.06 to 1.20)
Adjusted risk difference (95% CI)	+0.64% (0.22 to 1.14)	+1.73% (0.88 to 2.66)
Cardiovascular mortality		
Number of events/Total number of participants among diabetes group	164/3330 (4.9%)	282/1458 (19.3%)
Number of events/Total number of participants among mid-level HbA1c group	263/17162 (1.5%)	452/9254 (4.9%)
Adjusted risk ratio (95% CI)	1.05 (0.89 to 1.29)	1.20 (1.02 to 1.42)
Adjusted risk difference (95% CI)	+0.09% (-0.20 to 0.44)	+0.76% (0.08 to 1.53)

[a]200 iterations were performed for bootstrapping to estimate 95% confidence interval.

Table S4.5 Adjusted all-cause mortality risk ratio and risk difference at 5 and 10 years among participants with low HbA1c levels (4.0 to <5.0%) compared to those with mid-level HbA1c (5.0 to <5.7%) stratified by age and sex using parametric g-formula with pooled logistic regression models. [a,b]

	Follow-up periods	
	5 years	**10 years**
A) Age <65 years		
Number of events/Total number of participants among low HbA1c group	66/3183 (2.1%)	111/1988 (5.6%)
Number of events/Total number of participants among mid-level HbA1c group	330/14194 (2.3%)	610/8045 (7.7%)
Adjusted risk ratio (95% CI)	1.26 (0.99 to 1.59)	1.16 (0.96 to 1.34)
Adjusted risk difference (95% CI)	+0.60% (-0.03 to 1.28)	+0.80% (-0.21 to 1.72)
B) Age ≥65 years		
Number of events/Total number of participants among low HbA1c group	83/270 (30.7%)	138/217 (63.6%)
Number of events/Total number of participants among mid-level HbA1c group	795/3830 (20.8%)	1488/2861 (52.0%)
Adjusted risk ratio (95% CI)	1.19 (1.00 to 1.39)	1.05 (0.93 to 1.19)
Adjusted risk difference (95% CI)	+3.44% (-0.05 to 6.91)	+1.98% (-2.69 to 7.51)
C) Male		
Number of events/Total number of participants among low HbA1c group	79/1385 (5.7%)	136/836 (16.3%)
Number of events/Total number of participants among mid-level HbA1c group	678/8721 (7.8%)	1196/5342 (22.4%)
Adjusted risk ratio (95% CI)	1.11 (0.88 to 1.32)	1.07 (0.96 to 1.22)
Adjusted risk difference (95% CI)	+0.87% (-0.97 to 2.41)	+1.18% (-0.67 to 3.34)
D) Female		
Number of events/Total number of participants among low HbA1c group	70/2068 (3.4%)	113/1369 (8.3%)
Number of events/Total number of participants among mid-level HbA1c group	447/9303 (4.8%)	908/5564 (16.3%)
Adjusted risk ratio (95% CI)	1.61 (1.30 to 1.90)	1.21 (1.04 to 1.40)
Adjusted risk difference (95% CI)	+2.97% (1.59 to 4.32)	+2.41% (0.46 to 4.48)

[a]200 iterations were performed for bootstrapping to estimate 95% confidence interval.

[b]P-value for heterogeneity across sex (male vs female) for all-cause mortality at 5 years was 0.01 (p-value for heterogeneity in risk ratio) and 0.06 (p-value for heterogeneity in risk difference). P-values for heterogeneity for other estimates were >0.10

Table S4.6 Adjusted cardiovascular mortality risk ratio and risk difference at 5 and 10 years among participants with low HbA1c levels (4.0 to <5.0%) compared to those with mid-level HbA1c (5.0 to <5.7%) stratified by age and sex using parametric g-formula with pooled logistic regression models. [a,b]

	Follow-up periods	
	5 years	**10 years**
A) Age <65 years		
Number of events/Total number of participants among low HbA1c group	6/3123 (0.2%)	12/1889 (0.6%)
Number of events/Total number of participants among mid-level HbA1c group	39/13903 (0.3%)	85/7514 (1.1%)
Adjusted risk ratio (95% CI)	1.13 (0.52 to 2.18)	1.08 (0.60 to 1.64)
Adjusted risk difference (95% CI)	+0.05% (-0.19 to 0.41)	+0.08% (-0.38 to 0.57)
B) Age ≥65 years		
Number of events/Total number of participants among low HbA1c group	17/204 (8.3%)	34/113 (30.1%)
Number of events/Total number of participants among mid-level HbA1c group	224/3259 (6.9%)	367/1740 (21.1%)
Adjusted risk ratio (95% CI)	1.15 (0.78 to 1.73)	1.29 (0.93 to 1.80)
Adjusted risk difference (95% CI)	+0.92% (-1.25 to 4.09)	+3.66% (-0.92 to 10.0)
C) Male		
Number of events/Total number of participants among low HbA1c group	14/1320 (1.1%)	29/729 (4.0%)
Number of events/Total number of participants among mid-level HbA1c group	155/8198 (1.9%)	261/4407 (5.9%)
Adjusted risk ratio (95% CI)	1.24 (0.79 to 1.82)	1.36 (0.91 to 1.82)
Adjusted risk difference (95% CI)	+0.51% (-0.49 to 1.62)	+1.69% (-0.44 to 3.55)
D) Female		
Number of events/Total number of participants among low HbA1c group	9/2007 (0.5%)	17/1273 (1.3%)
Number of events/Total number of participants among mid-level HbA1c group	108/8964 (1.2%)	191/4847 (3.9%)
Adjusted risk ratio (95% CI)	1.22 (0.61 to 1.91)	1.18 (0.72 to 1.65)
Adjusted risk difference (95% CI)	+0.29% (-0.56 to 1.15)	+0.52% (-0.88 to 1.86)

[a]200 iterations were performed for bootstrapping to estimate 95% confidence interval.
[b] P-values for heterogeneity across age (<65 vs ≥65) and sex (male vs female) were >0.10 for all estimates.

Table S4.7 Adjusted all-cause mortality risk ratio and risk difference at 5 and 10 years among participants with low HbA1c levels (4.0 to <5.0%) compared to those with mid-level HbA1c (5.0 to <5.7%) using parametric g-formula with pooled logistic regression models, complete-case analysis (N=28,312). [a]

	Follow-up periods	
	5 years	**10 years**
Number of events/Total number of participants among low HbA1c group	78/2528 (3.1%)	137/1543 (8.9%)
Number of events/Total number of participants among mid-level HbA1c group	565/13149 (4.3%)	1134/7537 (15.0%)
Adjusted risk ratio (95% CI)	1.52 (1.19 to 1.82)	1.27 (1.06 to 1.49)
Adjusted risk difference (95% CI)	+2.33% (0.94 to 3.68)	+2.80% (0.61 to 5.25)

[a]200 iterations were performed for bootstrapping to estimate 95% confidence interval.

Table S4.8 Adjusted all-cause mortality risk ratio and risk difference at 5 and 10 years among participants with low HbA1c levels (4.0 to <5.0%) compared to those with mid-level HbA1c (5.0 to <5.7%) using parametric g-formula with pooled logistic regression models, restricting participants to those without anemia (N=34,740). [a,b]

	Follow-up periods	
	5 years	**10 years**
Number of events/Total number of participants among low HbA1c group	86/2993 (2.9%)	163/1877 (8.7%)
Number of events/Total number of participants among mid-level HbA1c group	875/16319 (5.4%)	1727/9828 (17.6%)
Adjusted risk ratio (95% CI)	1.25 (1.03 to 1.46)	1.08 (0.96 to 1.20)
Adjusted risk difference (95% CI)	+1.31% (0.16 to 2.40)	+1.01% (-0.49 to 2.34)

[a]200 iterations were performed for bootstrapping to estimate 95% confidence interval.

CHAPTER V

Conclusion and Public Health Implications

This dissertation examines the causal mechanisms involving diabetes and CVD using several causal modeling approaches. In the first study, using causal mediation analysis, I found that diabetes mediated around 10% of the pathway from low physical activity to all-cause mortality and cardiovascular events among older Mexican Americans. In the second study, adjusting for time-varying confounders with a marginal structural model, I found that diabetes and subsequent depressive symptoms were jointly associated with cardiovascular mortality among older Mexican Americans. These findings highlight the importance of diabetes prevention for physically inactive people and mental health management of diabetes among older Mexican Americans. Given the increasing prevalence of diabetes over the last century associated with significant increasing deaths and healthcare costs, this dissertation provides helpful information to build effective prevention and management strategies for diabetes among older Mexican Americans, a large racial/ethnic group with a high prevalence of diabetes.

In the third study, using several machine learning algorithms within the parametric g-formula, I found that low HbA1c levels, as well as diabetes, were associated with all-cause mortality among the U.S. general population. The findings have important clinical and public health implications that conventional diabetes care focusing on only high HbA1c might not be sufficient as low HbA1c also requires careful attention. Given that causal inference and machine learning methods have received great interest from not only epidemiologists/statisticians but also clinicians and policymakers, I hope this dissertation will be a guide for researchers to apply some causal inference and statistical methods to answer clinically important research questions.

REFERENCES

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1736-1788. doi:10.1016/S0140-6736(18)32203-7

2. Lobelo F, Rohm Young D, Sallis R, et al. Routine Assessment and Promotion of Physical Activity in Healthcare Settings: A Scientific Statement From the American Heart Association. *Circulation*. 2018;137(18):e495-e522. doi:10.1161/CIR.0000000000000559

3. Arem H, Moore SC, Patel A, et al. Leisure time physical activity and mortality: a detailed pooled analysis of the dose-response relationship. *JAMA Intern Med*. 2015;175(6):959-967. doi:10.1001/jamainternmed.2015.0533

4. Burchfiel CM, Sharp DS, Curb JD, et al. Physical activity and incidence of diabetes: the Honolulu Heart Program. *Am J Epidemiol*. 1995;141(4):360-368.

5. Monterrosa AE, Haffner SM, Stern MP, Hazuda HP. Sex difference in lifestyle factors predictive of diabetes in Mexican-Americans. *Diabetes Care*. 1995;18(4):448-456.

6. Reddigan JI, Ardern CI, Riddell MC, Kuk JL. Relation of physical activity to cardiovascular disease mortality and the influence of cardiometabolic risk factors. *Am J Cardiol*. 2011;108(10):1426-1431. doi:10.1016/j.amjcard.2011.07.005

7. Hamer M, Stamatakis E. Physical activity and risk of cardiovascular disease events: inflammatory and metabolic mechanisms. *Med Sci Sports Exerc*. 2009;41(6):1206-1211. doi:10.1249/MSS.0b013e3181971247

8. Centers for Disease Control and Prevention. Status Report for Step It Up! The Surgeon General's Call to Action to Promote Walking and Walkable Communities. 2017.

9. Keadle SK, McKinnon R, Graubard BI, Troiano RP. Prevalence and trends in physical activity among older adults in the United States: A comparison across three national surveys. *Prev Med*. 2016;89:37-43. doi:10.1016/j.ypmed.2016.05.009

10. US Department of Health and Human Services, Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2020. Atlanta, GA, 2020.

11. American Diabetes Association. 5. Lifestyle Management: Standards of Medical Care in Diabetes-2019. *Diabetes Care*. 2019;42(Suppl 1):S46-S60. doi:10.2337/dc19-S005

12. Brawley LR, Rejeski WJ, King AC. Promoting physical activity for older adults: the challenges for changing behavior. *Am J Prev Med*. 2003;25(3 Suppl 2):172-183.

13. Xu G, Liu B, Sun Y, et al. Prevalence of diagnosed type 1 and type 2 diabetes among US adults in 2016 and 2017: population based study. *BMJ*. 2018;362:k1497. doi:10.1136/bmj.k1497

14. Golden SH, Brown A, Cauley JA, et al. Health disparities in endocrine disorders: biological, clinical, and nonclinical factors--an Endocrine Society scientific statement. *J Clin Endocrinol Metab*. 2012;97(9):E1579-1639. doi:10.1210/jc.2012-2043

15. Spanakis EK, Golden SH. Race/ethnic difference in diabetes and diabetic complications. *Curr Diab Rep*. 2013;13(6):814-823. doi:10.1007/s11892-013-0421-9

16. Nordahl H, Lange T, Osler M, et al. Education and cause-specific mortality: the mediating role of differential exposure and vulnerability to behavioral risk factors. *Epidemiology*. 2014;25(3):389-396. doi:10.1097/EDE.0000000000000080

17. VanderWeele TJ. Mediation Analysis: A Practitioner's Guide. *Annu Rev Public Health*. 2016;37:17-32. doi:10.1146/annurev-publhealth-032315-021402

18. Bertone-Johnson ER, Tworoger SS, Hankinson SE. Recreational physical activity and steroid hormone levels in postmenopausal women. *Am J Epidemiol*. 2009;170(9):1095-1104. doi:10.1093/aje/kwp254

19. Shiels MS, Rohrmann S, Menke A, et al. Association of cigarette smoking, alcohol consumption, and physical activity with sex steroid hormone levels in US men. *Cancer Causes Control*. 2009;20(6):877-886. doi:10.1007/s10552-009-9318-y

20. Haan MN, Mungas DM, Gonzalez HM, Ortiz TA, Acharya A, Jagust WJ. Prevalence of dementia in older latinos: the influence of type 2 diabetes mellitus, stroke and genetic factors. *J Am Geriatr Soc*. 2003;51(2):169-177.

21. Ainsworth BE, Haskell WL, Whitt MC, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc*. 2000;32(9 Suppl):S498-504.

22. Piercy KL, Troiano RP, Ballard RM, et al. The Physical Activity Guidelines for Americans. *JAMA*. 2018;320(19):2020-2028. doi:10.1001/jama.2018.14854

23. Shih I-F, Haan MN, Paul KC, Yu Y, Sinsheimer JS, Ritz B. The Role of Physical Activity and Inflammation for Mortality, Cognition, and Depressive Symptoms in older Mexican Americans. *Am J Epidemiol*. 2019;[Epub ahead of print]. doi:10.1093/aje/kwz180

24. Mayeda ER, Haan MN, Kanaya AM, Yaffe K, Neuhaus J. Type 2 diabetes and 10-year risk of dementia and cognitive impairment among older Mexican Americans. *Diabetes Care*. 2013;36(9):2600-2606. doi:10.2337/dc12-2158

25. Wu JH, Haan MN, Liang J, Ghosh D, Gonzalez HM, Herman WH. Diabetes as a predictor of change in functional status among older Mexican Americans: a population-based cohort study. *Diabetes Care*. 2003;26(2):314-319. doi:10.2337/diacare.26.2.314

26. Aiello AE, Haan MN, Pierce CM, Simanek AM, Liang J. Persistent infection, inflammation, and functional impairment in older Latinos. *J Gerontol A Biol Sci Med Sci*. 2008;63(6):610-618. doi:10.1093/gerona/63.6.610

27. González HM, Haan MN, Hinton L. Acculturation and the prevalence of depression in older Mexican Americans: baseline results of the Sacramento Area Latino Study on Aging. *J Am Geriatr Soc*. 2001;49(7):948-953. doi:10.1046/j.1532-5415.2001.49186.x

28. James PA, Oparil S, Carter BL, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA*. 2014;311(5):507-520. doi:10.1001/jama.2013.284427

29. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. *Stat Med*. 1991;10(4):585-598.

30. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*. 1999;94(446):496-509.

31. Lau B, Cole SR, Gange SJ. Competing Risk Regression Models for Epidemiologic Data. *Am J Epidemiol*. 2009;170(2):244-256. doi:10.1093/aje/kwp107

32. Pearl J. The causal mediation formula--a guide to the assessment of pathways and mechanisms. *Prev Sci*. 2012;13(4):426-436. doi:10.1007/s11121-011-0270-1

33. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol*. 2013;42(5):1511-1519. doi:10.1093/ije/dyt127

34. Jeon CY, Lokken RP, Hu FB, van Dam RM. Physical activity of moderate intensity and risk of type 2 diabetes: a systematic review. *Diabetes Care*. 2007;30(3):744-752. doi:10.2337/dc06-1842

35. Elhakeem A, Cooper R, Whincup P, Brage S, Kuh D, Hardy R. Physical Activity, Sedentary Time, and Cardiovascular Disease Biomarkers at Age 60 to 64 Years. *J Am Heart Assoc*. 2018;7(16):e007459. doi:10.1161/JAHA.117.007459

36. WHO | Physical Activity and Older Adults. WHO. Accessed November 28, 2019. https://www.who.int/dietphysicalactivity/factsheet_olderadults/en/

37. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*. 2002;346(6):393-403. doi:10.1056/NEJMoa012512

38. Franco OH, de Laet C, Peeters A, Jonker J, Mackenbach J, Nusselder W. Effects of physical activity on life expectancy with cardiovascular disease. *Arch Intern Med*. 2005;165(20):2355-2360. doi:10.1001/archinte.165.20.2355

39. Berlin JA, Colditz GA. A meta-analysis of physical activity in the prevention of coronary heart disease. *Am J Epidemiol*. 1990;132(4):612-628.

40. Koolhaas CM, Dhana K, Golubic R, et al. Physical Activity Types and Coronary Heart Disease Risk in Middle-Aged and Elderly Persons: The Rotterdam Study. *Am J Epidemiol*. 2016;183(8):729-738. doi:10.1093/aje/kwv244

41. VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology*. 2014;25(4):473-484. doi:10.1097/EDE.0000000000000105

42. Mozaffarian D, Hao T, Rimm EB, Willett WC, Hu FB. Changes in diet and lifestyle and long-term weight gain in women and men. *N Engl J Med*. 2011;364(25):2392-2404. doi:10.1056/NEJMoa1014296

43. Rothenbacher D, Hoffmeister A, Brenner H, Koenig W. Physical activity, coronary heart disease, and inflammatory response. *Arch Intern Med*. 2003;163(10):1200-1205. doi:10.1001/archinte.163.10.1200

44. Borghouts LB, Keizer HA. Exercise and insulin sensitivity: a review. *Int J Sports Med*. 2000;21(1):1-12. doi:10.1055/s-2000-8847

45. Short KR, Vittone JL, Bigelow ML, et al. Impact of aerobic exercise training on age-related changes in insulin sensitivity and muscle oxidative capacity. *Diabetes*. 2003;52(8):1888-1896.

46. Röckl KSC, Witczak CA, Goodyear LJ. Signaling mechanisms in skeletal muscle: acute responses and chronic adaptations to exercise. *IUBMB Life*. 2008;60(3):145-153. doi:10.1002/iub.21

47. Holtermann A, Krause N, van der Beek AJ, Straker L. The physical activity paradox: six reasons why occupational physical activity (OPA) does not confer the cardiovascular health benefits that leisure time physical activity does. *Br J Sports Med*. 2018;52(3):149-150. doi:10.1136/bjsports-2017-097965

48. Kautzky-Willer A, Harreiter J, Pacini G. Sex and Gender Differences in Risk, Pathophysiology and Complications of Type 2 Diabetes Mellitus. *Endocr Rev*. 2016;37(3):278-316. doi:10.1210/er.2015-1137

49. Aune D, Norat T, Leitzmann M, Tonstad S, Vatten LJ. Physical activity and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis. *Eur J Epidemiol*. 2015;30(7):529-542. doi:10.1007/s10654-015-0056-z

50. Tobias M. Global control of diabetes: information for action. *Lancet*. 2011;378(9785):3-4. doi:10.1016/S0140-6736(11)60604-1

51. Logue J, Walker JJ, Colhoun HM, et al. Do men develop type 2 diabetes at lower body mass indices than women? *Diabetologia*. 2011;54(12):3003-3006. doi:10.1007/s00125-011-2313-3

52. Steen J, Loeys T, Moerkerke B, Vansteelandt S. Flexible Mediation Analysis With Multiple Mediators. *Am J Epidemiol*. 2017;186(2):184-193. doi:10.1093/aje/kwx051

53. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14(3):300-306.

54. Rodriguez CJ, Allison M, Daviglus ML, et al. Status of Cardiovascular Disease and Stroke in Hispanics/Latinos in the United States. *Circulation*. 2014;130(7):593-625. doi:10.1161/CIR.0000000000000071

55. CDC. Hispanic Health | VitalSigns. Centers for Disease Control and Prevention. Published May 5, 2015. Accessed June 25, 2020. https://www.cdc.gov/vitalsigns/hispanic-health/index.html

56. Harris MI. Racial and ethnic differences in health care access and health outcomes for adults with type 2 diabetes. *Diabetes Care*. 2001;24(3):454-459. doi:10.2337/diacare.24.3.454

57. Moulton CD, Pickup JC, Ismail K. The link between depression and diabetes: the search for shared mechanisms. *Lancet Diabetes Endocrinol*. 2015;3(6):461-471. doi:10.1016/S2213-8587(15)00134-5

58. Hackett RA, Steptoe A. Type 2 diabetes mellitus and psychological stress - a modifiable risk factor. *Nat Rev Endocrinol*. 2017;13(9):547-560. doi:10.1038/nrendo.2017.64

59. Harshfield EL, Pennells L, Schwartz JE, et al. Association Between Depressive Symptoms and Incident Cardiovascular Diseases. *JAMA*. 2020;324(23):2396-2405. doi:10.1001/jama.2020.23068

60. Rajan S, McKee M, Rangarajan S, et al. Association of Symptoms of Depression With Cardiovascular Disease and Mortality in Low-, Middle-, and High-Income Countries. *JAMA Psychiatry*. 2020;77(10):1052-1063. doi:10.1001/jamapsychiatry.2020.1351

61. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. *Diabetes Care*. 2001;24(6):1069-1078. doi:10.2337/diacare.24.6.1069

62. Yusuf S, Hawken S, Ounpuu S, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet*. 2004;364(9438):937-952. doi:10.1016/S0140-6736(04)17018-9

63. Inoue K, Beekley J, Goto A, Jeon CY, Ritz BR. Depression and Cardiovascular Disease Events among Patients with Type 2 Diabetes: A Systematic Review and Meta-Analysis with Bias Analysis. *Journal of Diabetes and its Complications*. Published online September 2, 2020:107710. doi:10.1016/j.jdiacomp.2020.107710

64. Garcia ME, Lee A, Neuhaus J, Gonzalez H, To TM, Haan MN. Diabetes Mellitus as a Risk Factor for Development of Depressive Symptoms in a Population-Based Cohort of Older Mexican Americans. *J Am Geriatr Soc*. 2016;64(3):619-624. doi:10.1111/jgs.14019

65. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550-560. doi:10.1097/00001648-200009000-00011

66. Inoue K, Mayeda ER, Paul KC, et al. The Association of Physical Activity with Cardiovascular Events and Mortality Mediated by Diabetes in Older Mexican Americans. *Am J Epidemiol*. Published online May 8, 2020. doi:10.1093/aje/kwaa068

67. Lyness JM, Noel TK, Cox C, King DA, Conwell Y, Caine ED. Screening for depression in elderly primary care patients. A comparison of the Center for Epidemiologic Studies-Depression Scale and the Geriatric Depression Scale. *Arch Intern Med*. 1997;157(4):449-454.

68. Black SA, Goodwin JS, Markides KS. The association between chronic diseases and depressive symptomatology in older Mexican Americans. *J Gerontol A Biol Sci Med Sci*. 1998;53(3):M188-194. doi:10.1093/gerona/53a.3.m188

69. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*. 1977;1(3):385-401. doi:10.1177/014662167700100306

70. Odden MC, Covinsky KE, Neuhaus JM, Mayeda ER, Peralta CA, Haan MN. The Association of Blood Pressure and Mortality Differs by Self-reported Walking Speed in Older Latinos. *J Gerontol A Biol Sci Med Sci*. 2012;67(9):977-983. doi:10.1093/gerona/glr245

71. Karim ME, Gustafson P, Petkau J, et al. Marginal Structural Cox Models for Estimating the Association Between β-Interferon Exposure and Disease Progression in a Multiple Sclerosis Cohort. *Am J Epidemiol*. 2014;180(2):160-171. doi:10.1093/aje/kwu125

72. Cortes-Bergoderi M, Goel K, Murad MH, et al. Cardiovascular mortality in Hispanics compared to non-Hispanic whites: A systematic review and meta-analysis of the Hispanic paradox. *European Journal of Internal Medicine*. 2013;24(8):791-799. doi:10.1016/j.ejim.2013.09.003

73. Liao Y, Cooper RS, Cao G, Kaufman JS, Long AE, McGee DL. Mortality from coronary heart disease and cardiovascular disease among adult U.S. Hispanics: findings from the National Health Interview Survey (1986 to 1994). *J Am Coll Cardiol*. 1997;30(5):1200-1205. doi:10.1016/s0735-1097(97)00278-7

74. Swenson CJ, Trepka MJ, Rewers MJ, Scarbro S, Hiatt WR, Hamman RF. Cardiovascular disease mortality in Hispanics and non-Hispanic whites. *Am J Epidemiol*. 2002;156(10):919-928. doi:10.1093/aje/kwf140

75. Novak M, Mucsi I, Rhee CM, et al. Increased Risk of Incident Chronic Kidney Disease, Cardiovascular Disease, and Mortality in Patients With Diabetes With Comorbid Depression. *Diabetes Care*. 2016;39(11):1940-1947. doi:10.2337/dc16-0048

76. Scherrer JF, Garfield LD, Chrusciel T, et al. Increased risk of myocardial infarction in depressed patients with type 2 diabetes. *Diabetes Care*. 2011;34(8):1729-1734. doi:10.2337/dc11-0031

77. Pan A, Lucas M, Sun Q, et al. Increased mortality risk in women with depression and diabetes mellitus. *Arch Gen Psychiatry*. 2011;68(1):42-50. doi:10.1001/archgenpsychiatry.2010.176

78. Bot M, Pouwer F, Zuidersma M, van Melle JP, de Jonge P. Association of coexisting diabetes and depression with mortality after myocardial infarction. *Diabetes Care*. 2012;35(3):503-509. doi:10.2337/dc11-1749

79. Cummings DM, Kirian K, Howard G, et al. Consequences of Comorbidity of Elevated Stress and/or Depressive Symptoms and Incident Cardiovascular Outcomes in Diabetes: Results From the REasons for Geographic And Racial Differences in Stroke (REGARDS) Study. *Diabetes Care*. 2016;39(1):101-109. doi:10.2337/dc15-1174

80. Egede LE, Nietert PJ, Zheng D. Depression and all-cause and coronary heart disease mortality among adults with and without diabetes. *Diabetes Care*. 2005;28(6):1339-1345. doi:10.2337/diacare.28.6.1339

81. Black SA, Markides KS, Ray LA. Depression predicts increased incidence of adverse health outcomes in older Mexican Americans with type 2 diabetes. *Diabetes Care*. 2003;26(10):2822-2828. doi:10.2337/diacare.26.10.2822

82. Carney RM, Freedland KE. Depression and coronary heart disease. *Nat Rev Cardiol*. 2017;14(3):145-155. doi:10.1038/nrcardio.2016.181

83. Veith RC, Lewis N, Linares OA, et al. Sympathetic nervous system activity in major depression. Basal and desipramine-induced alterations in plasma norepinephrine kinetics. *Arch Gen Psychiatry*. 1994;51(5):411-422. doi:10.1001/archpsyc.1994.03950050071008

84. Carney RM, Freedland KE, Stein PK, Skala JA, Hoffman P, Jaffe AS. Change in heart rate and heart rate variability during treatment for depression in patients with coronary heart disease. *Psychosom Med*. 2000;62(5):639-647.

85. Miller AH, Maletic V, Raison CL. Inflammation and its discontents: the role of cytokines in the pathophysiology of major depression. *Biol Psychiatry*. 2009;65(9):732-741. doi:10.1016/j.biopsych.2008.11.029

86. Broadley AJM, Korszun A, Jones CJH, Frenneaux MP. Arterial endothelial function is impaired in treated depression. *Heart*. 2002;88(5):521-523. doi:10.1136/heart.88.5.521

87. Shimbo D, Child J, Davidson K, et al. Exaggerated serotonin-mediated platelet reactivity as a possible link in depression and acute coronary syndromes. *Am J Cardiol*. 2002;89(3):331-333.

88. Chrousos GP. The role of stress and the hypothalamic-pituitary-adrenal axis in the pathogenesis of the metabolic syndrome: neuro-endocrine and target tissue-related causes. *Int J Obes Relat Metab Disord*. 2000;24 Suppl 2:S50-55.

89. Brummett BH, Babyak MA, Siegler IC, Mark DB, Williams RB, Barefoot JC. Effect of smoking and sedentary behavior on the association between depressive symptoms and mortality from coronary heart disease. *Am J Cardiol*. 2003;92(5):529-532.

90. Carney RM, Freedland KE, Eisen SA, Rich MW, Jaffe AS. Major depression and medication adherence in elderly patients with coronary artery disease. *Health Psychol*. 1995;14(1):88-90.

91. Forbes A, Murrells T, Mulnier H, Sinclair AJ. Mean HbA1c, HbA1c variability, and mortality in people with diabetes aged 70 years and older: a retrospective cohort study. *Lancet Diabetes Endocrinol*. 2018;6(6):476-486. doi:10.1016/S2213-8587(18)30048-2

92. Currie CJ, Peters JR, Tynan A, et al. Survival as a function of HbA(1c) in people with type 2 diabetes: a retrospective cohort study. *Lancet*. 2010;375(9713):481-489. doi:10.1016/S0140-6736(09)61969-3

93. Li F-R, Zhang X-R, Zhong W-F, et al. Glycated Hemoglobin and All-Cause and Cause-Specific Mortality Among Adults With and Without Diabetes. *J Clin Endocrinol Metab*. 2019;104(8):3345-3354. doi:10.1210/jc.2018-02536

94. Saydah S, Tao M, Imperatore G, Gregg E. GHb level and subsequent mortality among adults in the U.S. *Diabetes Care*. 2009;32(8):1440-1446. doi:10.2337/dc09-0117

95. Goto A, Arah OA, Goto M, Terauchi Y, Noda M. Severe hypoglycaemia and cardiovascular disease: systematic review and meta-analysis with bias analysis. *BMJ*. 2013;347:f4533. doi:10.1136/bmj.f4533

96. Palta P, Huang ES, Kalyani RR, Golden SH, Yeh H-C. Hemoglobin A1c and Mortality in Older Adults With and Without Diabetes: Results From the National Health and Nutrition Examination Surveys (1988-2011). *Diabetes Care*. 2017;40(4):453-460. doi:10.2337/dci16-0042

97. Selvin E, Steffes MW, Zhu H, et al. Glycated Hemoglobin, Diabetes, and Cardiovascular Risk in Nondiabetic Adults. *New England Journal of Medicine*. 2010;362(9):800-811. doi:10.1056/NEJMoa0908359

98. Carson AP, Fox CS, McGuire DK, et al. Low hemoglobin A1c and risk of all-cause mortality among US adults without diabetes. *Circ Cardiovasc Qual Outcomes*. 2010;3(6):661-667. doi:10.1161/CIRCOUTCOMES.110.957936

99. Paprott R, Schaffrath Rosario A, Busch MA, et al. Association between hemoglobin A1c and all-cause mortality: results of the mortality follow-up of the German National Health Interview and Examination Survey 1998. *Diabetes Care*. 2015;38(2):249-256. doi:10.2337/dc14-1787

100. Brewer N, Wright CS, Travier N, et al. A New Zealand linkage study examining the associations between A1C concentration and mortality. *Diabetes Care*. 2008;31(6):1144-1149. doi:10.2337/dc07-2374

101. Sakurai M, Saitoh S, Miura K, et al. HbA1c and the risks for all-cause and cardiovascular mortality in the general Japanese population: NIPPON DATA90. *Diabetes Care*. 2013;36(11):3759-3765. doi:10.2337/dc12-2412

102. Schöttker B, Rathmann W, Herder C, et al. HbA1c levels in non-diabetic older adults - No J-shaped associations with primary cardiovascular events, cardiovascular and all-cause mortality after adjustment for confounders in a meta-analysis of individual participant data from six cohort studies. *BMC Med*. 2016;14:26. doi:10.1186/s12916-016-0570-1

103. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet*. 2016;387(10027):1513-1530. doi:10.1016/S0140-6736(16)00618-8

104. Austin PC. Using Ensemble-Based Methods for Directly Estimating Causal Effects: An Investigation of Tree-Based G-Computation. *Multivariate Behavioral Research*. 2012;47(1):115-135. doi:10.1080/00273171.2012.640600

105. Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. *Stat Med*. 2010;29(3):337-346. doi:10.1002/sim.3782

106. Schuler MS, Rose S. Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies. *Am J Epidemiol*. 2017;185(1):65-73. doi:10.1093/aje/kww165

107. Snowden JM, Rose S, Mortimer KM. Implementation of G-Computation on a Simulated Data Set: Demonstration of a Causal Inference Technique. *Am J Epidemiol*. 2011;173(7):731-738. doi:10.1093/aje/kwq472

108. Kuhn M, Johnson K. *Applied Predictive Modeling*. Vol 26. Springer; 2013. Accessed January 11, 2020. http://appliedpredictivemodeling.com

109. National Center for Health Statistics, Centers for Disease Control and Prevention National Health and Nutrition Examination Survey. Published January 29, 2019. Accessed February 11, 2019. www.cdc.gov/nchs/nhanes.htm

110. National Center for Health Statistics, Centers for Disease Control and Prevention. NHANES response rates and population totals. Accessed February 18, 2019. https://wwwn.cdc.gov/nchs/nhanes/ResponseRates.aspx

111. National Center for Health Statistics, Centers for Disease Control and Prevention NCHS research ethics review board (ERB) approval. Published December 1, 2017. Accessed February 18, 2019. www.cdc.gov/nchs/ nhanes/irba98.htm

112. Christman AL, Lazo M, Clark JM, Selvin E. Low Glycated Hemoglobin and Liver Disease in the U.S. Population. *Diabetes Care*. 2011;34(12):2548-2550. doi:10.2337/dc11-0944

113. Selvin E, Parrinello CM, Sacks DB, Coresh J. Trends in prevalence and control of diabetes in the United States, 1988-1994 and 1999-2010. *Ann Intern Med*. 2014;160(8):517-525. doi:10.7326/M13-2411

114. Aggarwal V, Schneider ALC, Selvin E. Low Hemoglobin A1c in Nondiabetic Adults: An elevated risk state? *Diabetes Care*. Published online July 31, 2012. doi:10.2337/dc11-2531

115. NHANES MEC Laboratory Procedures Manual. Accessed October 19, 2019. https://wwwn.cdc.gov/nchs/data/nhanes/2013-2014/manuals/2013_MEC_Laboratory_Procedures_Manual.pdf

116. Fillenbaum GG, Burchett BM, Blazer DG. Identifying a national death index match. *Am J Epidemiol*. 2009;170(4):515-518. doi:10.1093/aje/kwp155

117. Inoue K, Ritz B, Brent GA, Ebrahimi R, Rhee CM, Leung AM. Association of Subclinical Hypothyroidism and Cardiovascular Disease With Mortality. *JAMA Netw Open*. 2020;3(2):e1920745. doi:10.1001/jamanetworkopen.2019.20745

118. Stekhoven DJ, Buhlmann P. MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112-118. doi:10.1093/bioinformatics/btr597

119. ranger: A Fast Implementation of Random Forests version 0.11.2 from CRAN. Accessed October 19, 2019. https://rdrr.io/cran/ranger/

120. xgboost: Extreme Gradient Boosting version 0.90.0.2 from CRAN. Accessed October 19, 2019. https://rdrr.io/cran/xgboost/

121. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol*. 2007;6:Article25. doi:10.2202/1544-6115.1309

122. Dickerman BA, Giovannucci E, Pernar CH, Mucci LA, Hernán MA. Guideline-Based Physical Activity and Survival Among US Men With Nonmetastatic Prostate Cancer. *Am J Epidemiol*. 2019;188(3):579-586. doi:10.1093/aje/kwy261

123. Keil AP, Edwards JK, Richardson DR, Naimi AI, Cole SR. The parametric G-formula for time-to-event data: towards intuition with a worked example. *Epidemiology*. 2014;25(6):889-897. doi:10.1097/EDE.0000000000000160

124. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ*. 2003;326(7382):219. doi:10.1136/bmj.326.7382.219

125. NHANES - Continuous NHANES Web Tutorial - Specifying Weighting Parameters. Accessed February 18, 2019. https://www.cdc.gov/nchs/tutorials/NHANES/SurveyDesign/Weighting/intro.htm

126. World Health Organization. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Published 2011. Accessed March 14, 2020. https://www.who.int/vmnis/indicators/haemoglobin.pdf

127. Kim C, Bullard KM, Herman WH, Beckles GL. Association Between Iron Deficiency and A1C Levels Among Adults Without Diabetes in the National Health and Nutrition

Examination Survey, 1999–2006. *Diabetes Care*. 2010;33(4):780-785. doi:10.2337/dc09-0836

128. American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. *Diabetes Care*. 2020;43(Suppl 1):S14-S31. doi:10.2337/dc20-S002

129. Davis IC, Ahmadizadeh I, Randell J, Younk L, Davis SN. Understanding the impact of hypoglycemia on the cardiovascular system. *Expert Rev Endocrinol Metab*. 2017;12(1):21-33. doi:10.1080/17446651.2017.1275960

130. Varlamov O, Bethea CL, Roberts CT. Sex-specific differences in lipid and glucose metabolism. *Front Endocrinol (Lausanne)*. 2014;5:241. doi:10.3389/fendo.2014.00241

131. Campbell PT, Newton CC, Patel AV, Jacobs EJ, Gapstur SM. Diabetes and cause-specific mortality in a prospective cohort of one million U.S. adults. *Diabetes Care*. 2012;35(9):1835-1844. doi:10.2337/dc12-0002

132. Rawshani A, Rawshani A, Franzén S, et al. Mortality and Cardiovascular Disease in Type 1 and Type 2 Diabetes. *New England Journal of Medicine*. 2017;376(15):1407-1418. doi:10.1056/NEJMoa1608664

133. Gregg EW, Cheng YJ, Srinivasan M, et al. Trends in cause-specific mortality among adults with and without diagnosed diabetes in the USA: an epidemiological analysis of linked national survey and vital statistics data. *The Lancet*. 2018;391(10138):2430-2440. doi:10.1016/S0140-6736(18)30314-3