Survival Analysis of the Effects of Health Insurance Status, Race, and Income on Diabetes Incidence

Lauren Bastian

Submitted under the supervision of Birgit Grund to the University Honors Program at the University of Minnesota-Twin Cities in partial fulfillment of the requirements for the degree of Bachelor of Arts, summa cum laude, in Statistical Practice.

September 2022

Introduction

An estimated 11.3% of the adult population of the United States has been diagnosed with diabetes, a proportion that has steadily increased over time (National Diabetes Statistics Report | Diabetes | CDC, 2022). Diabetes represents a significant personal and public health crisis, in some cases resulting in disability and premature death (Beckles & Thompson-Reid, 2001). Diabetes is more prevalent during middle age; among those aged 45-65, 14.5% have been diagnosed with diabetes. Women represent 46% of all diagnosed diabetes cases in the U.S., warranting investigation of factors for diabetes incidence for aging women (Beckles & Thompson-Reid, 2001, National Diabetes Statistics Report | Diabetes | CDC, 2022).

Diabetes poses an particular health risk for Black, Indigenous, and People of Color (BIPOC) in the United States, with American-Indian, Alaskan Native, non-Hispanic Black, Hispanic, and Asian Americans experiencing higher rates of diabetes than White, non-Hispanic Americans (National Diabetes Statistics Report | Diabetes | CDC, 2022). Racial health disparities may be partially explained by additional social stressors among people of color which result in earlier health deterioration among BIPOC than their white counterparts (Geronimus et al, 2006; Chyu & Upchurch, 2011; Williams, 2012). BIPOC women may experience negative confounding health effects as a result of dual social stressors, with Black/African-American women experiencing significantly earlier health deterioration than Black/African-American men and among women of other racial minorities (Geronimus et al, 2006; Chyu & Upchurch). BIPOC women are at higher risk for onset of diabetes even after adjusting for physical and behavioral characteristics. In a prospective study of race and diabetes incidence among middle-aged women, Black, Hispanic, and Asian participants had 136%, 114%, and 43% higher diabetes incidence, over 10 years than whites, respectively. After adjusting for physical and behavioral characteristics, the hazard ratio increased for Asian women and decreased for Black women, indicating that, although the prevalence of diabetes in Asian American women is similar to White women, they may have the greatest inherent risk for diabetes (Ma et al., 2012).

Lack of health insurance poses a serious health risk for Americans, with study by Wilper et al suggesting that after adjusting for race, income, health status, smoking, alcohol-use, exercise, and BMI, adult Americans without health insurance were at a higher risk for death than uninsured Americans (2009). Insurance status varies by household income and in 2018, an estimated 9.3% of adults aged 45-64 were uninsured. Households with lower income are significantly more likely to be uninsured, and significantly less likely to have private health insurance than their higher income counterparts (Berchick, 2019). By race, non-Hispanic whites are least likely to be uninsured, followed by Asian, black, and Hispanic Americans (Berchick, 2019). Those without health insurance are less likely to receive adequate health services with regards to diabetes management (Doucette et al., 2017), but although BIPOC and low income persons disproportionately lack health insurance, there has been little research on the effects of race, income, and health insurance status on diabetes incidence.

Prevalence of diabetes in the United States has increased along with obesity rates (Weir & Jan, 2022). The World Health Organization defines class I, II, and III obesity as BMI between 30 and 34.99, BMI between 35 and 39.99, and BMI above 40 respectively (Weir & Jan, 2022). Obese (classes I and II) and severely obese individuals are at elevated risk for developing heart disease and type II diabetes than non-obese individuals (Narayan et al., 2007, Weir & Jan, 2022). Race, income, and many other social determinants of health confound the relationship between obesity and diabetes, with BIPOC and lower income groups experiencing higher rates of both obesity and diabetes (Hill-Briggs et al., 2020).

The presence of censored time-to event data motivates the use of survival analysis methods to investigate the associations of health insurance status, socioeconomic status, race, and BMI on diabetes incidence. Survival analysis allows researchers to assess the effect of factors of interest on the risk for event incidence using time-to-event data, which quantifies the time between an initial event and a subsequent event of interest (Machin et al., 2006, Kleinbaum & Klein, 2005).

Time-to event data differs from other types of continuous, numerical data in that it allows varying follow-up times across individuals. For example, one individual may experience the event of interest in three years, where another may experience the event in ten years, and another may not never experience the event during the span of the study (Machin et al., 2006, Kleinbaum & Klein, 2005). Such unobserved events are considered "censored" and differ from missing data in that censored individuals are not removed or imputed from the data set as their inclusion offer additional insight about event risk (Machin et al., 2006, Kleinbaum & Klein, 2005). The most common form of censoring is right censoring. There are three primary reasons an individual may be right-censored. First, the individual does not experience the event by the end study. Second, the individual withdraws from the study. Third, the individual is lost to follow-up (Kleinbaum & Klein, 2005). Other types of censoring, including 'left-censoring' which occurs when the true survival time is shorter than or equal to the observed survival time due to a delay in diagnosis of event, and 'interval censoring' which occurs when the true event incidence is unknown but observed within a certain time interval, are less common (Radke, 2003, Kleinbaum & Klein, 2005). Proper handling of censored data allows for a larger analytical sample, as censored individuals are not removed from the risk set. That is, even if an individual never develops diabetes, they are still considered 'at risk' for incidence over the course of the study (Machin et al., 2006, George et al., 2014).

Definition of the study start time must be carefully considered. For clinical trials start time is often the time of treatment randomization, however risk exposure is less clear for observational health studies (Machin et al., 2006). For analysis of long-term longitudinal observational data it is often more intuitive to use age as the time scale rather than the study duration. However, if the individuals enter the study at different ages, individuals are often stratified by birth-cohort, to control for the environmental effects that may contribute to event risk (Korn et al., 1997, Hurley,

2015). For example, an average individual born in 1950 may have healthier eating habits than an average individual born in 1960, resulting in lower risk for diabetes onset.

The present paper first uses Kaplan-Meier curves to illustrate the relationship between health insurance status, race, and household income and diabetes risk among middle-aged American women, then estimates hazard ratios using the Cox proportional hazards model.

Methods

SWAN Dataset

To investigate the association between health insurance status, race, and household income and diabetes incidence among middle-aged American women, we apply survival analysis methods to the Study of Women's Health Across the Nation (SWAN) dataset. This longitudinal observational study aims to "help scientists, health care providers, and women learn how mid-life experiences affect health and quality of life during aging," among middle-aged American women living in designated geographic areas including Inkster, Michigan; Hackensack, New Jersey; Chicago, Illinois; Ypsilanti, Michigan; Los Angeles, California; Pittsburgh, Pennsylvania; and Boston, Massachusetts (Sutton-Tyrell et al., 2014). Women spoke at least one of English, Japanese, Cantonese, and Spanish, and belonged to one of five target racial/ethnic groups including African American, Chinese American, Japanese American, Hispanic American, and Non-Hispanic White American. Between the years of 1994 and 2008, data were collected on various demographic, physical, medical, social, economic, and emotional characteristics during a baseline visit and 10 follow-up visits at medical centers in the designated regions. Of the 3302 women recruited into the study, 2245 were retained for their tenth follow-up visit. Although the New Jersey site is included in the study, data were not collected from the New Jersey site during the tenth visit (Sutton-Tyrell et al., 2014). The study is ongoing and now includes 16 follow-up visits with the original cohort with visit 17 in progress at the time of the present paper.

The analytical sample includes 2686 women who participated in SWAN between the years 1994 and 2008, did not have diabetes at baseline, and had records for age at all visits, health insurance status at baseline, race, household income at baseline, BMI at baseline, and diabetes status for at least one follow-up visit.

Diabetes Incidence

Since diabetes status was not directly assessed during the baseline visit, baseline diabetes status was evaluated as whether the individual had been prescribed insulin by their doctor or other healthcare provider in the last month. Those

who were taking insulin at baseline were not included in the survival dataset since the event of interest had already occurred prior to study entry.

For visits 1 to 10, diabetes status was assessed with the question, "Since your last study visit, has a doctor, nurse practitioner or other health care provider told you that you had any of the following conditions or treated you for them?" and followed by a list of medical conditions, including diabetes. Possible responses included "Yes", "No", and "Don't Know".

From the analytical sample we evaluate time to diabetes occurrence. We use a chronological time-scale based on age rather than visit number, where time at baseline $t = 0$ is the individual's age during the baseline study, because visits cover a two year period and the chronological time-scale enhances interpretability of results (Machin et al., 2006, Hurley, 2015). Incident diabetes was determined as the age at which the individual first reported a diabetes diagnosis since the last visit.

Covariates

Health Insurance Status

Health insurance status was assessed only during the baseline visit with the question "Which of the following categories best describes how you usually pay for your medical care?" with responses including prepaid private insurance, other private insurance, Medicare, Medicaid, military or veteran's administration-sponsored, no insurance, and other non-specified insurance (Sutton-Tyrell et al., 2010). Insurance status was binarized to evaluate differences in diabetes survival for insured and uninsured women. Individuals who reported having any of prepaid private insurance, other private insurance, Medicare, Medicaid, military or veteran's administration-sponsored, or other non-specified insurance were considered insured, while those who reported no insurance only were considered uninsured. Individuals who reported both no insurance and another form of insurance were considered of unknown insurance status and were excluded due to the very few response of this type.

Race

Race was assessed during a screening survey prior to the baseline study as African American, Chinese American, Japanese American, Hispanic American, or Non-Hispanic White American. Individuals were recorded as belonging to only one racial/ethnic group.

Baseline Household Income

Household income was assessed with the question, "What is your total family income (before taxes) from all

sources within your household in the last year?" with responses including "less than $19,999", "$20,000 to $49,999", "$50,000 to $99,999", "$100,000 or more", and "Don't know". Those who had no response or marked unknown were not included in the dataset.

Baseline BMI

All physical measures were taken by a medical professional. Body Mass Index (BMI) was calculated as weight in kilograms divided by the square height in meters by a health. For comparison of obesity levels, continuous BMI is categorized into three groups: Not obese (BMI < 30), obese ($30 \leq$ BMI < 40), and severely obese (BMI ≥ 40). The obese group includes class I and class II obesity and the severely obese group includes class III obesity as defined by the National Institute of Health and World Health Organization (Weir & Jan, 2022).

Survival Models

Kaplan-Meier Curve

Survival curves illustrate the incidence of diabetes over time, by plotting the cumulative proportion of participants who have not yet experienced the event over follow-up time in the study. Kaplan-Meier (K-M) curves are a non-parametric method of estimating a survival function $S(t)$, the probability that a random individual's survival time T exceeds a specific time t (Kleinbaum & Klein, 2005, Machin et al., 2006). K-M estimates $S(t)$ based on observed survival times and a risk set for each time t, defined as the number of individuals who have survived or remain uncensored up to t. The estimated K-M survival curve is defined by a step-wise function (1)

$$\begin{aligned} \hat{S}(t_j) &= \prod_{i=1}^{j} \hat{\Pr}(T > t_i | T \geq t_i) \\ &= \hat{S}(t_{j-1}) \times \hat{\Pr}(T > t_j | T \geq t_j) \end{aligned} \tag{1}$$

where $\hat{S}(t_{j-1})$ is the probability of an individual surviving beyond a time t_{j-1}, defined by (2) times the conditional probability of surviving beyond time t given that the individual has survived up to time t (Kleinbaum & Klein, 2005, Machin et al., 2006).

$$\hat{S}(t_{j-1}) = \prod_{i=1}^{j-1} \hat{\Pr}(T > t_i | T \geq t_i) \tag{2}$$

K-M curves for multiple groups are compared using the log-rank test, which compares observed versus compares

observed versus expected distributions of event time using a large sample chi-squared test. The test compares the null hypothesis that the distribution of event times are equal across all groups versus the alternative that event incidence between groups varies at least one time point during the study. The log-rank test statistic is commonly approximated by (3).

$$X^2 = \sum_{i=1}^{G} \frac{(O_i - E_i)^2}{E_i} \tag{3}$$

Under H_0 the test statistic approximately follows a chi-squared distribution with $G - 1$ degree of freedom, where G is the number of groups/survival curves being compared, O_i are observed event counts for each group $i \in 1, G$ and E_i are the expected event counts for each group $i \in 1, G$ (Kleinbaum & Klein, 2005).

Cox Proportional Hazards Model

The Cox proportional hazards model (CPH) is a semi-parametric model used to estimate a hazard function adjusted for several covariates. CPH estimates a hazard function of the form

$$h(t, \mathbf{X}) = h_0(t) e^{\sum_{i=1}^{p} \beta_i X_i} \tag{4}$$

where $\mathbf{X}$ is the predictors $X_i \ldots X_p$, h_0 is the baseline hazard function, and β are the estimated coefficients for each of the predictors. The baseline hazard function is time-dependent while the predictors $\mathbf{X}$ are assumed to be time-independent to satisfy the proportional hazards assumption (Kleinbaum & Klein, 2005). In this study, we assume that health insurance status, household income, and BMI are time-independent such that baseline measures are indicative of these measures across all visits of the study. To formally assess the proportional hazards assumption for CPH, we use a a Goodness-of-Fit test based on the Schoenfeld residuals (Grambsch & Therneau, 1994). The proportional hazards assumption was met for all models with $p > 0.05$ for all Goodness-of-Fit tests.

Covariates that do not satisfy the proportional hazards assumption but must be controlled in the model are stratified. Adjusting by strata allows adjustment by such covariates by assuming separate baseline hazard functions for each strata. From adjusted CPH models, hazard ratios can be obtained for covariates that satisfy the proportional hazards assumption, but not for the strata (Kleinbaum & Klein, 2005).

Stratified CPH models further assume no interaction between the estimated coefficients $\hat{\beta}$ and the strata; while baseline hazard functions differ between strata $\hat{\beta}$ are equivalent across strata (Kleinbaum & Klein, 2005).

In this study, models were stratified by age by creating three age groups defined by age upon study entry equal to 42-45 ($n_{42-45} = 1275$), 46-49 ($n_{46-49} = 1117$), and 50-53 ($n_{50-53} = 295$). Provided the sample size of each age group is sufficiently large, stratification results in no loss of parameter estimation precision versus adjusting by age as a confounding covariate in the CPH model, and is a well established method of adjustment to account for unknown environmental effects of birth year (Korn et al., 1997, Hurley, 2015). A Likelihood Ratio Test is used to assess the no-interaction assumption, by using a Chi-Squared Test of Deviance (with degrees of freedom equal to the difference in the number of estimated coefficients) to compare the stratified CPH model including an interaction between the strata and the covariates and the stratified CPH model excluding the interaction between the strata and the covariates (Kleinbaum & Klein, 2005). In this study, with $p > 0.05$ for all models, all stratified CPH models satisfy the no-interaction assumption. No significant differences are found between the estimated coefficients for each strata.

Hazard Ratio

The effects of covariates on diabetes incidence is quantified using the hazard ratio (HR), defined by (5), where $\hat{h}(t, \mathbf{X}^*)$ is the estimated hazard function for one individual with covariates $\mathbf{X}^*$ and $\hat{h}(t, \mathbf{X})$ is the estimated hazard function for another individual with covariates $\mathbf{X}$ (Klein & Kleinbaum, 2005).

$$\hat{\text{HR}} = \frac{\hat{h}(t, \mathbf{X}^*)}{\hat{h}(t, \mathbf{X})} \tag{5}$$

$\hat{\text{HR}} > 1$ indicates that an individual with covariates $\mathbf{X}^*$ is at greater risk for the event at time t than an individual with covariates $\mathbf{X}$. Conversely, $\hat{\text{HR}} < 1$ indicates that the individual with covariates $\mathbf{X}^*$ is at lower risk for the event at time t than an individual with covariates $\mathbf{X}$. $\hat{\text{HR}} = 1$ indicates no difference in the risk for the event (Klein & Kleinbaum, 2005).

In this study, we evaluate the effect of insurance status, race, BMI, and income, versus several reference levels, defining the denominator of the hazard ratio by these reference levels. The reference level for race is Non-Hispanic White, as this group has been shown to be at lower risk for incident diabetes; the reference level for insurance status is uninsured; the reference level for BMI group is BMI ¡ 30 lower, as lower BMI is associated with lower risk for incident diabetes; and the reference level for income is $50,000-$99,000 per year as this was the median and mode income level reported at baseline.

Results

Descriptive Statistics for Demographics and Incident Diabetes

In this study of a racially and socioeconomically diverse sample of middle-aged women who participated in the SWAN Study, a ten-year longitudinal, observational study of aging women ($n = 2686$).

Table 1: Table #: BMI Counts and Proportions by Race

Race	**BMI <30 n (%)**	**30 ≤ BMI <40 n (%)**	**BMI ≥ 40 n (%)**	**Row Totals n (%)**
Non-Hispanic White	939 (70.13%)	317 (23.67%)	83 (6.20%)	1339 (49.85%)
Black/African American	339 (50.15%)	253 (37.43%)	84 (12.43%)	676 (25.17%
Chinese/Chinese American	224 (96.14%)	7 (3.00%)	2 (0.86%)	233 (8.67%)
Japanese/Japanese American	231 (95.06%)	11 (4.53%)	1 (0.41%)	243 (9.05%)
Hispanic	127 (65.13%)	52 (26.67%)	16 (8.21%)	195 (7.26%)
Column Totals	1860 (69.25%)	640 (23.83%)	186 (6.92%)	2686

Note:

Table 2: Table #: Incident Diabetes by Race

Race	**Incident Diabetes within Racial Group n (%)**
Non-Hispanic White	101 (8.16%)
Black/African American	115 (20.50%)
Chinese/Chinese American	19 (8.88%)
Japanese/Japanese American	19 (8.48%)
Hispanic	20 (65.13%)
Column Totals	274 (10.20%)

Note:

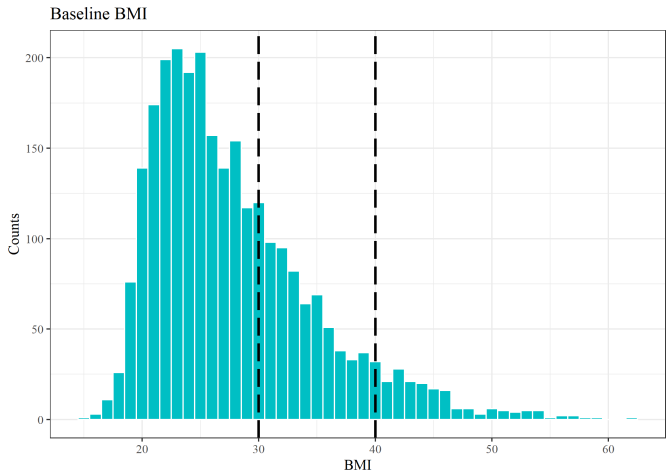

Figure 1: Figure #: Baseline BMI for all women in the analytical sample ($n = 2686$)

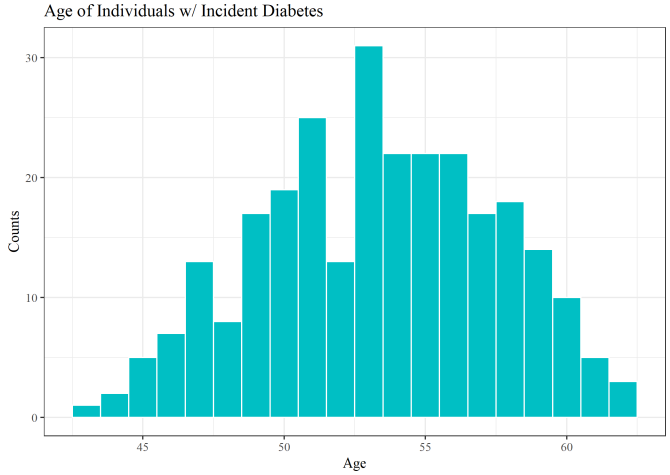

Figure 2: Figure #: Age of women with incident diabetes ($n = 274$)

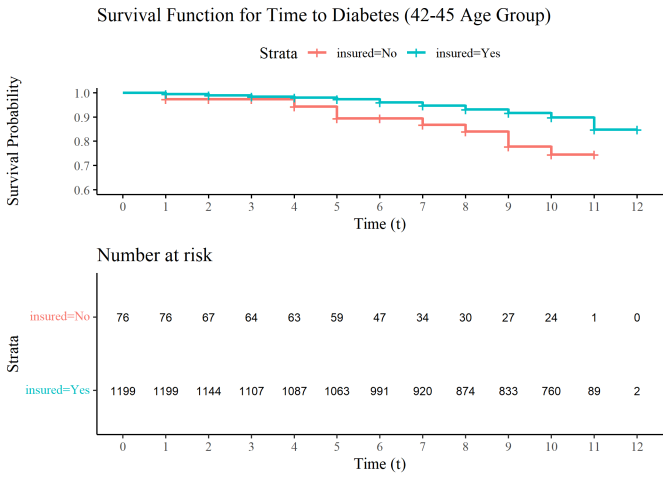

Figure 3: Figure #: Kaplan-Meier curve for diabetes incidence by insurance status among women aged 42-45 ($n = 1275$)

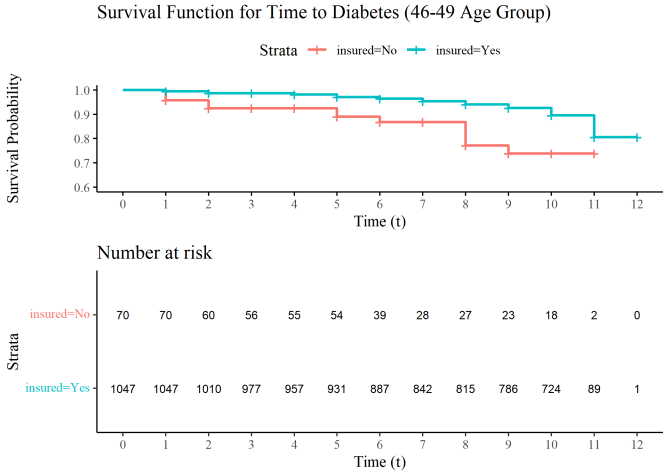

Figure 4: Figure #: Kaplan-Meier curve for diabetes incidence by insurance status among women aged 46-49 ($n = 1117$)

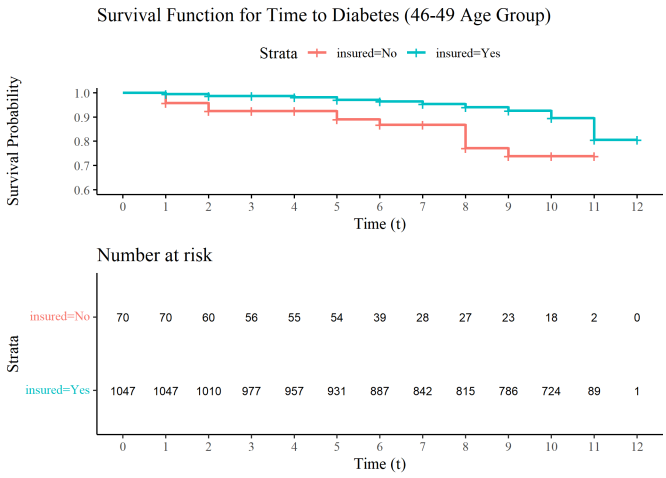

Figure 5: Figure #: Kaplan-Meier curve for diabetes incidence by insurance status among women aged 46-49 ($n = 1117$)

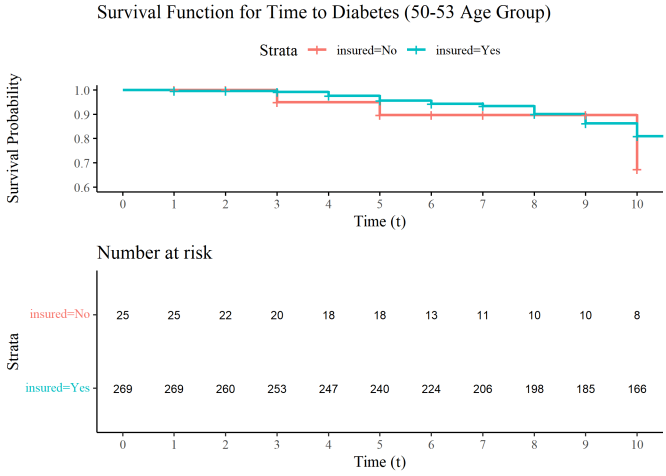

Figure 6: Figure #: Kaplan-Meier curve for diabetes incidence by insurance status among women aged 50-53 ($n = 294$)

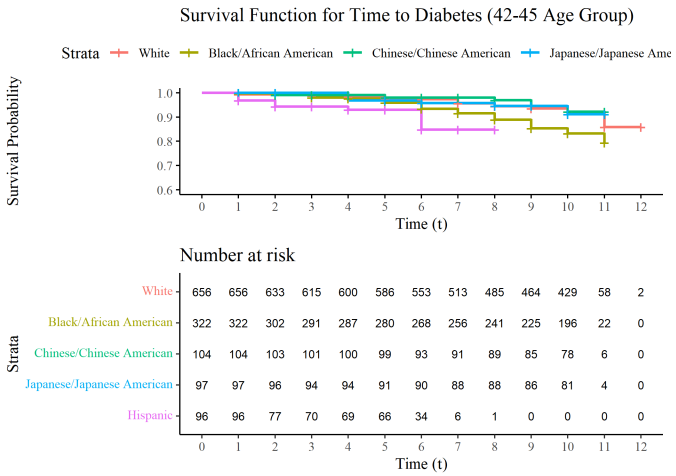

Figure 7: Figure #: Kaplan-Meier curve for diabetes incidence by race among women aged 42-45 ($n = 1275$)

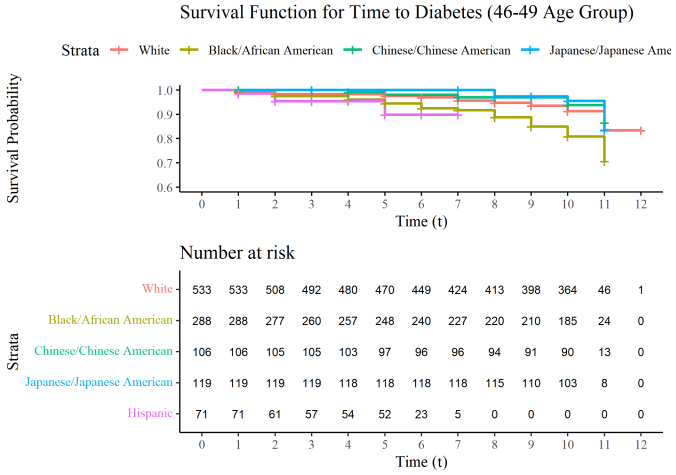

Figure 8: Figure #: Kaplan-Meier curve for diabetes incidence by race among women aged 46-49 ($n = 1117$)

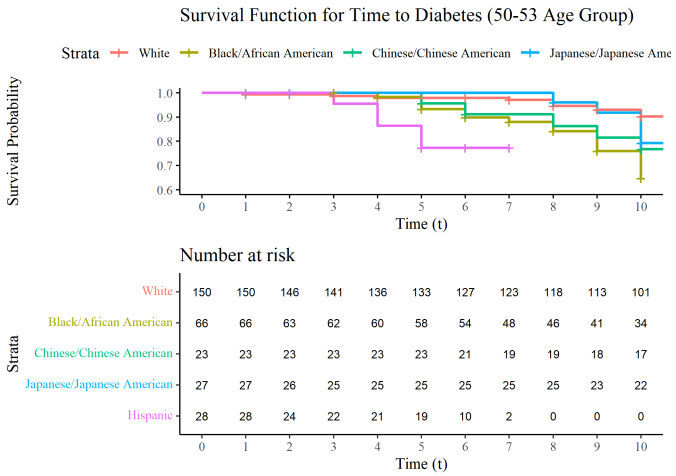

Figure 9: Figure #: Kaplan-Meier curve for diabetes incidence by race among women aged 50-53 ($n = 294$)

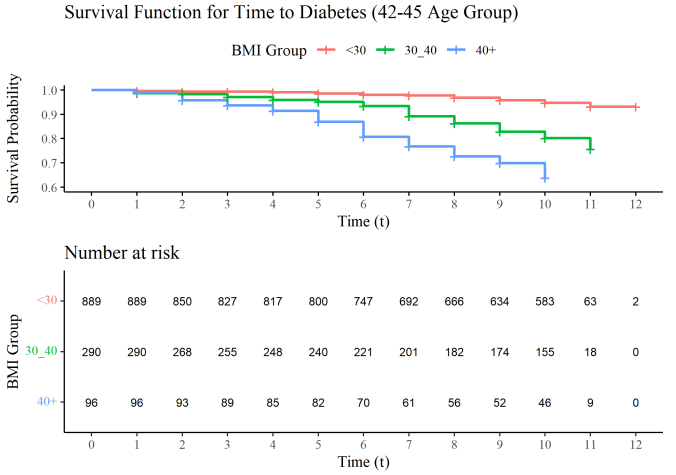

Figure 10: Figure #: Kaplan-Meier curve for diabetes incidence by BMI among women aged 42-45 ($n = 1275$)

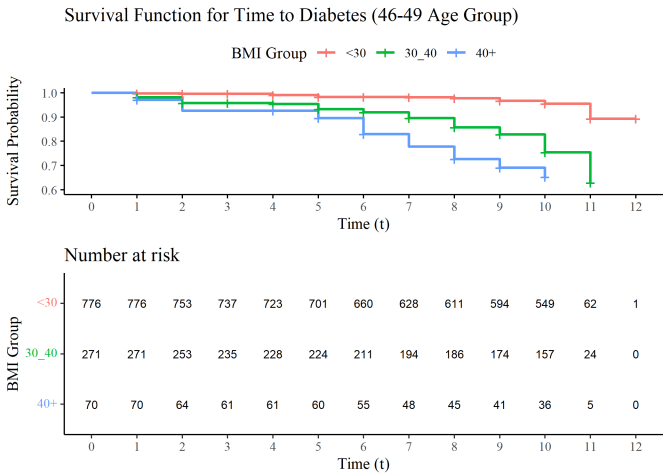

Figure 11: Figure #: Kaplan-Meier curve for diabetes incidence by BMI among women aged 46-49 ($n = 1117$)

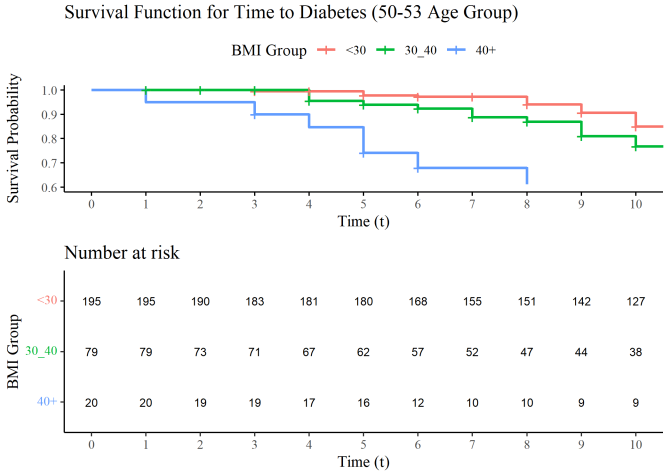

Figure 12: Figure #: Kaplan-Meier curve for diabetes incidence by BMI among women aged 50-53 ($n = 294$)

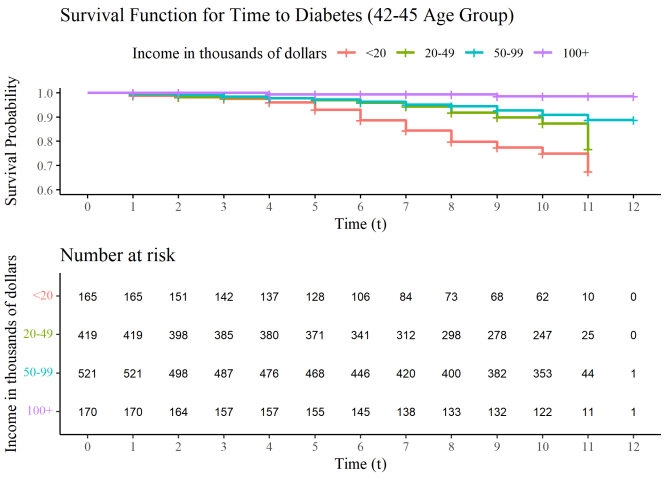

Figure 13: Figure #: Kaplan-Meier curve for diabetes incidence by income among women aged 42-45 ($n = 1275$)

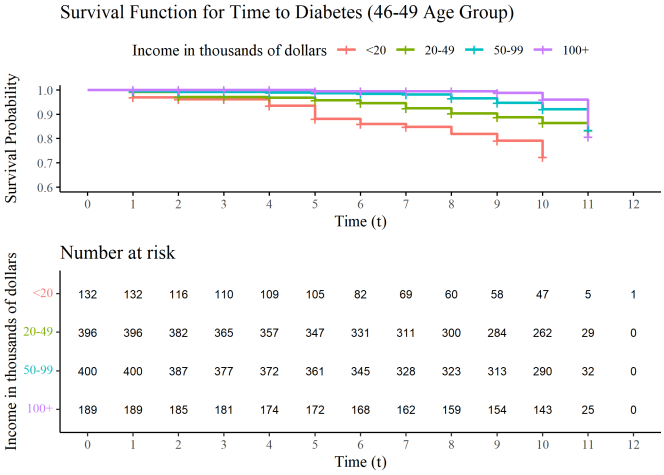

Figure 14: Figure #: Kaplan-Meier curve for diabetes incidence by income among women aged 46-49 ($n = 1117$)

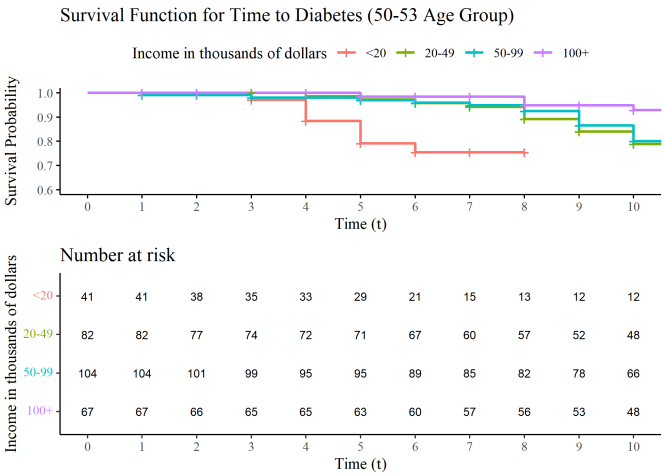

Figure 15: Figure #: Kaplan-Meier curve for diabetes incidence by income among women aged 50-53 ($n = 294$)

Discussion

References

Beckles, G. L., & Thompson-Reid, P. E. (2001). *Diabetes and women's health across life stages: A public health perspective*. Department of Health, Human Services, Centers for Disease Control, Prevention, National Center for Chronic Disease Prevention, and Health Promotion, Division of Diabetes Translation. Atlanta, GA, United States. https://webharvest.gov/peth04/20041107113818/http://www.cdc.gov/diabetes/pubs/pdf/women.pdf

Berchick, E. (2019, September). *Health insurance coverage in the united states: 2018*. U.S. Department of Commerce. Washington, D.C. https://heinonline-org.ezp3.lib.umn.edu/HOL/Page?handle=hein.beal/hltiscv0001&id=1&collection=usfed&index=

Chyu, L., & Upchurch, D. M. (2011). Racial and ethnic patterns of allostatic load among adult women in the united states: Findings from the national health and nutrition examination survey 1999–2004. *Journal of Women's Health*, *20*(4), 575–583. https://doi.org/10.1089/jwh.2010.2170

Colen, C. G., Ramey, D. M., Cooksey, E. C., & Williams, D. R. (2018). Racial disparities in health among nonpoor african americans and hispanics: The role of acute and chronic discrimination. *Social Science & Medicine*, *199*, 167–180. https://doi.org/10.1016/j.socscimed.2017.04.051

Doucette, E. D., Salas, J., Wang, J., & Scherrer, J. F. (2017). Insurance coverage and diabetes quality indicators among patients with diabetes in the US general population. *Primary Care Diabetes*, *11*(6), 515–521. https://doi.org/10.1016/j.pcd.2017.05.007

Evans, J. M. M., Newton, R. W., Ruta, D. A., MacDonald, T. M., & Morris, A. D. (2000). Socio-economic status, obesity and prevalence of type 1 and type 2 diabetes mellitus. *Diabetic Medicine*, *17*(6), 478–480. https://doi.org/10.1046/j.1464-5491.2000.00309.x

George, B., Seals, S., & Aban, I. (2014). Survival analysis and regression models. *Journal of Nuclear Cardiology*, *21*(4), 686–694. https://doi.org/10.1007/s12350-014-9908-2

Geronimus, A. T., Hicken, M., Keene, D., & Bound, J. (2006). "weathering" and age patterns of allostatic load scores among blacks and whites in the united states. *American Journal of Public Health*, *96*(5), 826–833. https://doi.org/10.2105/AJPH.2004.060749

Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, *81*(3), 515–526. https://doi.org/10.2307/2337123

Hill-Briggs, F., Adler, N. E., Berkowitz, S. A., Chin, M. H., Gary-Webb, T. L., Navas-Acien, A., Thornton, P. L., & Haire-Joshu, D. (2020). Social determinants of health and diabetes: A scientific review. *Diabetes Care*, *44*(1), 258–279. https://doi.org/10.2337/dci20-0053

Hurley, M. A. (2015). A reference relative time-scale as an alternative to chronological age for cohorts with long follow-up. *Emerging Themes in Epidemiology*, *12*(1), 18. https://doi.org/10.1186/s12982-015-0043-6

Kleinbaum, D. G., & Klein, M. (2005). *Survival analysis: A self-learning text* (2nd ed). Springer.

Lee, T. C., Glynn, R. J., Peña, J. M., Paynter, N. P., Conen, D., Ridker, P. M., Pradhan, A. D., Buring, J. E., & Albert, M. A. (2011). Socioeconomic status and incident type 2 diabetes mellitus: Data from the women's health study. *PLOS ONE*, *6*(12), e27670. https://doi.org/10.1371/journal.pone.0027670

Lidfeldt, J., Li, T. Y., Hu, F. B., Manson, J. E., & Kawachi, I. (2007). A prospective study of childhood and adult socioeconomic status and incidence of type 2 diabetes in women. *American Journal of Epidemiology*, *165*(8), 882–889. https://doi.org/10.1093/aje/kwk078

Ma, Y., Hébert, J. R., Manson, J. E., Balasubramanian, R., Liu, S., Lamonte, M. J., Bird, C. E., Ockene, J. K., Qiao, Y., Olendzki, B., Schneider, K. L., Rosal, M. C., Sepavich, D. M., Wactawski-Wende, J., Stefanick, M. L., Phillips, L. S., Ockene, I. S., Kaplan, R. C., Sarto, G. E., … Howard, B. V. (2012). Determinants of racial/ethnic disparities in incidence of diabetes in postmenopausal women in the u.s. *Diabetes Care*, *35*(11), 2226–2234. https://doi.org/10.2337/dc12-0412

Machin, D., Cheung, Y. B., Parmar, M. K. B., & Parmar, M. K. B. (2006). *Survival analysis: A practical approach* [OCLC: 68045541]. Wiley. Retrieved March 21, 2022, from http://www.123library.org/book_details/?id=57613

Narayan, K., Boyle, J. P., Thompson, T. J., Gregg, E. W., & Williamson, D. F. (2007). Effect of BMI on lifetime risk for diabetes in the u.s. *Diabetes Care*, *30*(6), 1562–1566. https://doi.org/10.2337/dc06-2544

National diabetes statistics report — diabetes — CDC. (2022, January 20). Retrieved March 20, 2022, from https://www.cdc.gov/diabetes/data/statistics-report/index.html

Piette, J. D., Wagner, T. H., Potter, M. B., & Schillinger, D. (2004). Health insurance status, cost-related medication underuse, and outcomes among diabetes patients in three systems of care. *Medical Care*, *42*(2), 102–109. https://doi.org/10.1097/01.mlr.0000108742.26446.17

Radke, B. R. (2003). A demonstration of interval-censored survival analysis. *Preventive Veterinary Medicine*, *59*(4), 241–256. https://doi.org/10.1016/S0167-5877(03)00103-X

Socioeconomic status of women with diabetes—united states, 2000. (2002). *JAMA*, *287*(19), 2496–2497. https://doi.org/10.1001/jama.287.19.2496

Starfield, B., Shi, L., & Macinko, J. (2005). Contribution of primary care to health systems and health. *The Milbank Quarterly*, *83*(3), 457–502. https://doi.org/10.1111/j.1468-0009.2005.00409.x

Sutton-Tyrell, K., Selzer, F., R. (Mary Francis Roy) Sowers, M., Neer, R., Powell, L., Gold, E. B., Greendale, G., Weiss, G., Matthews, K., & McKinlay, S. (2014). Study of women's health across the nation (SWAN), 2001-2003: Visit 05 dataset: Version 2 [Type: dataset]. https://doi.org/10.3886/ICPSR30501.V2

Sutton-Tyrell, K., Selzer, F., Sowers, M. R. (F. R., Finkelstein, J., Powell, L., Gold, E., Greendale, G., Weiss, G., Matthews, K., & Brooks, M. M. (2014a). Study of women's health across the nation (SWAN), 2004-2006: Visit 08 dataset: Version 2 [Type: dataset]. https://doi.org/10.3886/ICPSR32122.V2

Sutton-Tyrell, K., Selzer, F., Sowers, M. R. (F. R., Finkelstein, J., Powell, L., Gold, E., Greendale, G., Weiss, G., Matthews, K., & Brooks, M. M. (2014b). Study of women's health across the nation (SWAN), 2005-2007: Visit 09 dataset: Version 2 [Type: dataset]. https://doi.org/10.3886/ICPSR32721.V2

Sutton-Tyrell, K., Selzer, F., Sowers, M. R. (F. R., Finkelstein, J., Powell, L., Gold, E. B., Greendale, G., Weiss, G., & Matthews, K. (2014a). Study of women's health across the nation (SWAN), 2002-2004: Visit 06 dataset: Version 2 [Type: dataset]. https://doi.org/10.3886/ICPSR31181.V2

Sutton-Tyrell, K., Selzer, F., Sowers, M. R. (F. R., Finkelstein, J., Powell, L., Gold, E. B., Greendale, G., Weiss, G., & Matthews, K. (2014b). Study of women's health across the nation (SWAN), 2003-2005: Visit 07 dataset: Version 2 [Type: dataset]. https://doi.org/10.3886/ICPSR31901.V2

Sutton-Tyrrell, K., Selzer, F., Sowers, M., Finkelstein, J., Powell, L., Gold, E., David, G., Weiss, G., Matthews, K., & Brooks, M. M. (2014). Study of women's health across the nation (SWAN), 2006-2008: Visit 10 dataset: Version 2 [Type: dataset]. https://doi.org/10.3886/ICPSR32961.V2

Sutton-Tyrrell, K., Selzer, F., Sowers, M., R. (Mary Frances Roy), Neer, R., Powell, L., Gold, E. B., Greendale, G., Weiss, G., Matthews, K. A., & McKinlay, S. (2011a). Study of women's health across the nation (SWAN): Visit 01 dataset, [united states], 1997-1999: Version 3 [Type: dataset]. https://doi.org/10.3886/ICPSR29221.V3

Sutton-Tyrrell, K., Selzer, F., Sowers, M., R. (Mary Frances Roy), Neer, R., Powell, L., Gold, E. B., Greendale, G., Weiss, G., Matthews, K. A., & McKinlay, S. (2011b). Study of women's health across the nation (SWAN): Visit 02 dataset, [united states], 1998-2000: Version 4 [Type: dataset]. https://doi.org/10.3886/ICPSR29401.V4

Sutton-Tyrrell, K., Selzer, F., Sowers, M., R. (Mary Frances Roy), Neer, R., Powell, L., Gold, E. B., Greendale, G., Weiss, G., Matthews, K. A., & McKinlay, S. (2011c). Study of women's health across the nation (SWAN): Visit 03 dataset, [united states], 1999-2001: Version 4 [Type: dataset]. https://doi.org/10.3886/ICPSR29701.V4

Sutton-Tyrrell, K., Selzer, F., Sowers, M., R. (Mary Frances Roy), Neer, R., Powell, L., Gold, E. B., Greendale, G., Weiss, G., Matthews, K. A., & McKinlay, S. (2011d). Study of women's health across the nation (SWAN): Visit 04 dataset, [united states], 2000-2002: Version 4 [Type: dataset]. https://doi.org/10.3886/ICPSR30142.V4

Villarroel, M., & Cohen, R. (2016, May). *Health insurance continuity and health care access and utilization, 2014* (No. 249). National Center for Health Statistics. Hyattsville, MD. https://permanent.fdlp.gov/gpo69556/db249.pdf

Weir, C. B., & Jan, A. (2022). BMI classification percentile and cut off points. *StatPearls*. StatPearls Publishing. Retrieved April 26, 2022, from http://www.ncbi.nlm.nih.gov/books/NBK541070/

Williams, D. R. (2012). Miles to go before we sleep: Racial inequities in health. *Journal of Health and Social Behavior*, *53*(3), 279–295. Retrieved March 20, 2022, from https://www.jstor.org/stable/41725217

Wilper, A. P., Woolhandler, S., Lasser, K. E., McCormick, D., Bor, D. H., & Himmelstein, D. U. (2009). Health insurance and mortality in US adults. *American Journal of Public Health*, *99*(12), 2289–2295. https://doi.org/10.2105/AJPH.2008.157685

Appendix