# ESSAYS ON AGING-RELATED DISEASE IN THE U.S. POPULATION

Ezra Fishman

A DISSERTATION

in

Demography

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

_____

Irma T. Elo
Professor of Sociology

Co-Supervisor of Dissertation

_____

Samuel H. Preston
Professor of Sociology

Graduate Group Chairperson

_____

Michel Guillot
Associate Professor of Sociology

Dissertation Committee:

Douglas C. Ewbank, Research Professor Emeritus of Sociology

ACKNOWLEDGMENTS

Many people deserve thanks for contributing to this dissertation. Professor Sam Preston has been extraordinarily generous with his time and expertise. I am privileged to have studied under one of the great American demographers, and he has been a truly first-rate teacher and professional mentor to me. Professor Irma Elo, my committee chair, has also been very giving of her time and energies toward this thesis, especially the many iterations of Chapter 3. Professor Doug Ewbank has provided much guidance as well, especially for Chapter 2. I am lucky to have had such a helpful and conscientious a committee.

I would also like to thank Professors Jere Behrman, Warren Bilker, Michel Guillot, Kristen Harknett, Jason Karlawish, John MacDonald, and Dylan Small, for contributing to parts of this thesis and to my advancement toward the PhD more generally.

The following professionals have also been particularly helpful in various parts of the dissertation process: Tanya Yang, Dawn Ryan, Addie Metivier, Nykia Perez, Jim Kane, Chris Mustazza, Aron Buchman, Brenda Plassman, Kenneth Langa, Mary Beth Ofstedal, and Ryan MacKinnon.

My colleagues in the Penn Demography PhD program have made it a positive, collaborative experience. Andrew Stokes and Jamaica Corker have been important mentors to me. Thanks also to Nikkil Sudharsanan, Arun Hendi, Thomas Anderson, David Frankenfield, Megan Costa, Laura Kelly, and Edith Gutierrez.

My biggest thanks go to my parents, Lyle Fishman and Debra Rosenman, my brother Noam, and my wife Nicki Brodie. You have been unfailingly supportive and helpful.

Note: A version of Chapter 1 (Cohort Dynamics of Diabetes) was published as: Fishman E, Stokes A, and Preston SH. "The Dynamics of Diabetes among Birth Cohorts in the United States." *Diabetes Care*, 2014 Apr; 37(4):1052-9. doi: 10.2337/dc13-1982.

A version of Chapter 3 (Incident Diabetes and Mobility Limitations) was published as: Fishman E. "Incident Diabetes and Mobility Limitations: Reducing Bias through Risk-Set Matching." *Journals of Gerontology, Series A: Medical Sciences*, 2015 Jul; 70(7):860-5. doi: 10.1093/gerona/glu212.

## ABSTRACT

ESSAYS ON AGING-RELATED DISEASE IN THE U.S. POPULATION

Ezra Fishman

Dissertation supervised by Samuel H. Preston and Irma T. Elo

I research three topics in adult morbidity in the United States, focusing on two increasingly prevalent chronic diseases, diabetes and dementia. In the first essay, I investigate changes in age-specific diabetes prevalence across cohorts born in the 20th century and use the cross-cohort comparisons to generate model age patterns of diabetes prevalence and incidence. I show that most of the increase in diabetes prevalence over time is attributable to increases in age-specific prevalence from one cohort to the next. Because the risk of diabetes is embodied in cohorts, diabetes prevalence is likely to increase in the future even if the prevalence of risk factors such as obesity plateau.

In the second essay, I use multiple-decrement life tables to estimate age-specific lifetime risks of dementia for a dementia-free person. I estimate that about a quarter of dementia-free 70-year-old males and a third of females will develop dementia in their lifetimes. Although interventions that delay dementia onset could substantially reduce dementia risk, the results indicate a widespread need to prepare for a life stage with dementia.

In the third essay, I use recent advances in propensity-score matching techniques to estimate the association between incidence diabetes and subsequent accumulation of mobility limitations. Among observationally similar pairs of individuals, those who

developed diabetes reported an average of 25% more mobility limitations at study exit than those who did not develop diabetes. My estimates of this association are smaller than those found in most of the existing literature, but they are likely less biased.

## Contents

List of Tables

Main tables



Appendix Tables

List of Figures

Main figures



Appendix Figures

## Introduction

As adult mortality continues its long-run decline in industrialized countries, demographers have paid increasing attention to morbidity (Crimmins & Beltrán-Sánchez 2010; Vaupel 2010). In the following chapters, I research three topics in adult morbidity in the United States, focusing on two increasingly prevalent chronic diseases, diabetes and dementia. Increases in diabetes prevalence are associated with changes in the risk profile of the population, especially the prevalence and duration of obesity (Abdullah et al. 2011; Gregg, Boyle, et al. 2013; Reis et al. 2013). Increases in dementia prevalence are largely attributable to the declining risk of death at "younger old" ages, allowing a larger fraction individuals to live long enough to develop dementia (Deckers et al. 2014; Hurd, Martorell, Delavande, Mullen, & Langa 2013; Plassman et al. 2011).

A life-course perspective on chronic disease in the United States unites the three chapters that follow. In Chapter 1, I consider the life course of cohorts, using the fact that young-age characteristics of a cohort stay with the cohort as it ages, predicting later-life cohort characteristics. This perspective enables the estimation of model age patterns of diabetes prevalence and incidence that reflect what cohorts actually experience as they age. In Chapter 2, I investigate the life-cycle implications of dementia incidence and mortality for an average older adult. The life table methods used in Chapter 2 synthesize population rates back to the level of the individual life course. In Chapter 3, I isolate the role of incident diabetes in the subsequent accumulation of mobility limitations, reducing biases found in studies that use diabetes presence as the exposure.

Another unifying theme is the distinction – and interaction – between disease incidence and prevalence. Incidence is a classic demographic rate: the ratio of new cases to person-time at risk. Prevalence is the proportion of the population with the disease at a given time. Chapter 1 uses an innovative cohort-based method to infer incidence from prevalence and differential mortality. Chapter 2 uses incidence and prevalence to refine estimates of differential mortality, and synthesizes the different rates to develop the lifetime risk estimate. Lifetime risk would be equivalent to prevalence in a stationary population if there were no differential mortality; the large mortality differentials between people with dementia and their age-contemporaries without it necessitate a life table based estimate (Preston, Heuveline, & Guillot 2001). Finally, Chapter 3 shows that examining incident disease as a risk factor for subsequent outcomes, in this case mobility limitations, can generate somewhat different results than examining prevalent disease as a risk factor.

***The cohort dynamics of diabetes in the United States***

(*Note: This chapter was co-authored with Prof. Samuel H. Preston and Andrew Stokes.*)

Diabetes is a leading cause of mortality, disability, and health care costs in the United States (Murray 2013). Few national studies provide detail on age patterns of diabetes prevalence above age 65 (Y. J. Cheng et al. 2013; Menke, Casagrande, Geiss, & Cowie 2015). Since future gains in life expectancy are likely to be concentrated at old ages (Li, Lee, & Gerland 2013), such age-detail is valuable. Furthermore, data limitations reduce our ability to get good estimates of age-specific disease incidence, and of trends

over time, especially since changes in diagnostic criteria for diabetes and awareness of the disease among the public make self-reported data of questionable quality (e.g. Gregg et al. 2014). Most generally, there is insufficient understanding in chronic disease epidemiology of the importance of birth cohorts as embodiments of lifetimes of risk factors that presage future trends in prevalence and incidence.

Chapter 1 addresses these gaps in the literature by examining age patterns of diabetes prevalence across birth cohorts, using a stable measure of blood glucose collected from population-representative samples in the National Health and Nutrition Examination Surveys (NHANES), 1988 to 2010. In the cross-section, prevalence appears to decline after age 75, a result that could arise if people with diabetes at those ages die more quickly than new cases occur. However, Chapter 1 will show that this result appears in the cross-section because older cohorts had lower diabetes prevalence throughout their lives than younger cohorts had.

The value of the cohort-based approach is demonstrated in an age/period/cohort model of diabetes prevalence. To avoid the classic identification problem in such models, where any one of age, period, and cohort is a linear combination of the other two (Mason & Fienberg 1985), the chapter uses a single continuous measure to capture cohort-based risks of diabetes: the prevalence of obesity in the cohort at age 25. In addition to avoiding the classic identification problem, this approach also identifies cohort effects with an intuitively appealing risk factor, rather than with complex mathematical transformations (e.g. Reither et al. 2009). I find that including this single continuous variable reduces all period coefficients to nearly zero. In other words, increases in diabetes prevalence over

time are almost entirely explained by increases in the age-25 obesity prevalence of successively younger cohorts.

In a cohort closed to migration, increases in diabetes prevalence from one age to the next result only from incidence and mortality differentials between the diabetic and non-diabetic populations. Using the cohort perspective, I develop a model age pattern of diabetes incidence by measuring within-cohort changes in prevalence and adjusting for differential mortality. The results show that diabetes incidence peaks in late middle age (55 to 64) and declines at older ages. To my knowledge, these are the first estimates of diabetes incidence using nationally representative data with a stable, objective measure of blood glucose.

The main insight of Chapter 1 is that the risk of diabetes is embodied in cohorts. One implication is that the possible plateauing of obesity prevalence recently observed (Flegal, Carroll, Kit, & Ogden 2012) does not imply that diabetes prevalence will soon plateau as well. As current young-adult cohorts age, they are likely to experience higher levels diabetes prevalence at any given age than past cohorts. More generally, a perspective that integrates the life course of a cohort into the measurement of disease occurrence can provide deeper understandings of past trends and perhaps better ability to predict future trends.

***Lifetime risk of dementia in the United States***

Dementia imposes a financial cost of over $40,000 per affected person per year, comparable to the financial costs of heart disease and cancer (Hurd et al. 2013).

Americans over 60 reported fearing dementia more than any other disease, including cancer (Alzheimer's Association 2014). As the U.S. population ages, the number of Americans with dementia is very likely to increase in the coming decades (He & Larsen 2014; Kasper, Freedman, Spillman, & Wolff 2015). Using nationally representative, longitudinal data from the Aging, Demographics, and Memory Study (ADAMS), fielded 2001 to 2009, Chapter 2 estimates the probability that an average dementia-free person will develop dementia in the course of life.

An approach that incorporates the competing risks of death and dementia incidence in a prospective cohort allows one to estimate the risk that the average dementia-free individual will develop dementia in the future. It can also provide an estimate of the related measures of dementia-free life expectancy and life expectancy with dementia. These quantities are important for individuals, businesses, and governments as they plan for retirement, save and contribute to pensions, and assess future health care costs and caregiving needs. For demographers and epidemiologists, these quantities provide meaningful insight into the question of whether long-run gains in survivorship are being experienced in healthy or unhealthy states (Crimmins & Beltrán-Sánchez 2010; Crimmins, Hayward, & Hagedorn 2009).

To my knowledge, this is the first study of the life cycle implications of dementia that uses data representative of the U.S. old-age population and explicitly accounts for the competing risk of death. The study will also utilize stationary population relations to refine estimates of a difficult-to-estimate quantity, the age-specific relative risk of death with dementia (versus without dementia). Finally, I will use the competing-risks model to

simulate the effects of interventions that delay onset of dementia or reduce its risk throughout old age.

My findings suggest that the lifetime risk of dementia for an average adult are considerably higher than estimates extrapolated from non-national samples (Seshadri & Wolf 2007; Seshadri et al. 1997). Furthermore, my simulations suggest that, even in a population subject to interventions that substantially delay or reduce the risk of dementia, more than 20% of dementia-free 70-year-olds are likely to develop the disease before they die. The results suggest a widespread and underappreciated need to prepare for a life stage with dementia.

***Incident diabetes and mobility limitations: reducing bias using risk-set matching***

Although diabetes is an increasingly important cause of death in the U.S. (Murray 2013), its consequences in terms of morbidity are also important. In Chapter 3, I investigate the relation between incident diabetes and the accumulation of mobility limitations using the Health and Retirement Study (HRS), a longitudinal survey representative of the U.S. population above age 50, covering 1992 to 2010. There is no agreed-upon method for estimating functional limitations associated with chronic conditions in the presence of comorbidities. Although diabetes prevalence and physical functional limitations are strongly correlated, the level of rigor of the studies showing this correlation is fairly low (Wong & Gregg 2013). Most studies, for example, ignore the possibility of reverse causality: mobility limitations can increase the risk of diabetes by making persons more sedentary (Bardenheier, Gregg, Zhuo, Cheng, & Geiss 2014).

Studies looking at diabetes presence as a risk factor for mobility limitation also cannot control for the time-ordering of covariates, even though many personal characteristics can both cause and result from diabetes (Narayan et al. 2011). Recent advances in propensity-score matching techniques have not been applied to the study of chronic diseases.

Chapter 3 addresses these gaps in the literature by investigating the role of diabetes incidence, rather than prevalence, in subsequent mobility decline. The long duration and time-invariance of questions about diabetes presence and functional limitations in the HRS allows the investigator to observe both incident diabetes cases and a long subsequent accumulation of mobility limitations. The rich data on socioeconomic and health backgrounds of subjects also allow for more robust controls for confounding variables compared to most prior literature. The innovative method of risk-set matching non-parametrically controls for time-varying onset of diabetes and of time-varying and time-invariant covariates. Cases and controls not only look similar in the cross-section; they have similar pre-exposure *trajectories*. This procedure generates a stronger, less-biased control group than the extant literature. The chapter illustrates one approach that can be used to study the contribution of a specific disease to physical functioning limitations in the presence of multiple comorbidities.

The results show that individuals who developed diabetes subsequently accumulated more mobility limitations than matched controls. In each pair of case and control, subjects were followed for the same length of time both before and after diabetes onset. The magnitude of the association between diabetes and mobility limitations is smaller than that presented in most of the prior literature; however, there is reason to

believe my estimate is less biased. More practically, there is a great deal of room for better diabetes management to reduce the burden of physical functioning limitations associated with diabetes.

# 1 The Cohort Dynamics of Diabetes in the United States

## 1.1 INTRODUCTION

Diabetes is a leading cause of death in the United States (Murphy, Xu, & Kochanek 2013). A recent meta-analysis estimates that people with diabetes have a 50-80% increased risk of disability, including impaired mobility, activities of daily living, and instrumental activities of daily living, compared to people without diabetes (Wong et al. 2013). The prevalence of diabetes among adults is approximately 12%, corresponding to approximately 26.1 million adults with diabetes in 2005-10 (Y. J. Cheng et al. 2013).

The incidence and prevalence of type 2 diabetes, which accounts for over 90% of diabetes cases (American Diabetes Association 2012), are clearly related to factors in an individual's past. In particular, individuals' own histories of obesity and smoking (Luo et al. 2013; Yeh, Duncan, Schmidt, Wang, & Brancati 2010) have been shown to affect the risk of developing diabetes. Of these risk factors, the relationship between obesity history and diabetes incidence has been studied more extensively. One study found a steep gradient in the lifetime risk of diabetes based on body mass index (BMI, measured in kilograms per meters squared) at age 18. Males in the optimal BMI range of 18.5 to 25 $kg/m^2$ at age 18 had a 19.8% lifetime risk of diabetes, while males with BMI in the obese range of 30 to 35 $kg/m^2$ at age 18 had a 57.0% lifetime risk of diabetes (Narayan et al. 2007). A European cohort study found that the earlier in life that subjects gained weight, the more likely they were to develop diabetes (Schienkiewitz, Schulze, Hoffmann, Kroke, & Boeing 2006). Among subjects in the Framingham Heart Study, each additional two years of obesity were associated with about a 12% increased odds of developing diabetes

(Abdullah et al. 2011). In the National Longitudinal Study of Adolescent Health, persistent obesity was associated with twice the risk of diabetes prevalence compared to adult-onset obesity (The, Richardson, & Gordon-Larsen 2013). In the CARDIA study, each additional year a person was obese increased their odds of developing diabetes by 4% (Reis et al. 2013). These and other studies indicate that obesity over the life course is an important predictor of diabetes incidence.

In this paper, we investigate the rise in diabetes in the United States through the lens of birth cohorts. Previous studies examining changes in diabetes prevalence over time have compared one calendar-year period to another (Bullard et al. 2013; Y. J. Cheng et al. 2013). However, like other chronic diseases, type 2 diabetes is the result of cumulative processes that develop over a lifetime. A full understanding of the prevalence of diabetes at a moment in time requires reference to the past, a past that is embodied in the birth cohorts alive during that period. Because histories in a birth cohort are persistent – characteristics of a birth cohort established at age 25 remain the age-25 characteristics of that cohort as it ages – we expect to find "cohort effects" that differentiate one birth cohort from another as they age.

Birth cohorts not only embody a history of exposures, they are also the appropriate vehicle for calculating disease incidence. We take advantage of this opportunity to present new estimates of the age-pattern of diabetes incidence in the United States. These are the first estimates of incidence that use measured data in a nationally representative sample. Previous national estimates of diabetes incidence used

retrospective reports of individuals rather than biological indicators and provided little age detail (Geiss et al. 2006; Narayan et al. 2003).

## 1.2 METHODS

### *Population and data collection*

In order to investigate the dynamics of diabetes in the United States, we use data from the National Health and Nutrition Examination Surveys (NHANES). We employ data from NHANES III, conducted in two phases, 1988 to 1991 and 1991 to 1994; and from the Continuous NHANES that began in 1999, for which data are released in two-year cycles. We pool adjacent data-release cycles of Continuous NHANES to obtain three observation periods from Continuous NHANES: 1999 to 2002, 2003 to 2006, and 2007 to 2010. NHANES is a complex, multi-stage probability sample of the U.S. civilian non-institutionalized population. Participants complete a home interview and are then examined in a mobile examination center, which includes sampling participants' blood for laboratory tests. Participants are randomized into morning or afternoon examinations, and the morning examinees are asked to fast for at least nine hours prior to the examination. Whenever possible, NHANES uses consistent laboratory procedures over time to facilitate analysis of trends in population health. The National Center for Health Statistics (NCHS) provides extensive documentation of NHANES survey, examination, and laboratory procedures on its website (National Health and Nutrition Examination Survey 2012). The characteristics of the NHANES study sample are reported elsewhere (Bullard et al. 2013; Y. J. Cheng et al. 2013).

There were 88,224 individuals examined during our study periods. We exclude individuals below age 20 (n=40,899), above age 80 (n=3,558), or who were pregnant (n=1,510). We also exclude individuals who were exactly 20 years old when surveyed in 2010 (n=105) because these individuals would not comprise a complete birth cohort, as described below. We also exclude subjects with missing HbA1c values (n=2,022). The final analytic sample for HbA1c-based measures consists of 40,130 observations, with 7,011 observations from Phase 1 of NHANES III, 7,427 from Phase 2 of NHANES III, 7,778 from NHANES 1999-2002, 7,755 from NHANES 2003-2006, and 10,159 from NHANES 2007-2010.

***Definition of diabetes***

We rely on laboratory results, rather than self-reported diagnoses, because the latter fails to capture the considerable number of individuals in the US population with undiagnosed diabetes. A 2010 study estimated that 3.9 million individuals above age 20 had undiagnosed diabetes, representing 19% of the diabetic population (Cowie et al. 2010). Furthermore, intertemporal comparisons based on self-reported diagnosis are complicated by the fact that criteria for diagnosing diabetes in the clinical setting have changed (Gregg et al. 2004; Stokes & Mehta 2013).

Our primary definition of diabetes is based on HbA1C, which was first measured in NHANES III. This measure reflects average glycemia over a prolonged period and thus has more intra-subject stability than the leading alternative, a measure of fasting plasma glucose (FPG) (Bonora & Tuomilehto 2011). Furthermore, HbA1c-based

measures of diabetes are more strongly associated with cardiovascular disease and death than are FPG-based measures (Selvin et al. 2010). Finally, only 54% as many observations of diabetes status are available in NHANES using FPG as using HbA1c. A sensitivity analysis defined diabetes presence as FPG ≥ 126 mg/dL**.**

Several changes in laboratory measurement of HbA1C occurred over the course of Continuous NHANES (detailed elsewhere (Bullard et al. 2013)), but we follow the NCHS recommendation and the methods of recent studies and used HbA1C data without any corrections or adjustments (Bullard et al. 2013; Y. J. Cheng et al. 2013). Individuals are considered diabetic if they had HbA1c ≥ 6.5% (48 mmol/mol) (American Diabetes Association 2012). Because diabetes medication is expected to reduce glycemia, the HbA1c values of medicated persons might not capture their diabetes status correctly; therefore, all individuals who reported taking diabetes medication are considered diabetic. In our sample, there were 4,678 individuals who met our definition of having diabetes. There were 896 individuals, or 19.2% of the group with diabetes, who reported taking diabetes medication and who had HbA1c < 6.5%.

***Cohort assignment***

Birth cohorts must be constructed from repeated cross-sections because NHANES does not repeatedly sample the same individuals over time. We calculate each individual's birth year using the equation *Birth cohort = Period - Age*. For the purpose of calculating birth cohorts, *Period* is defined as the midpoint of the NHANES wave or phase: April 21, 1990 for Phase 1 of NHANES III, April 23, 1992 for Phase 2 of

NHANES III, and January 1 of the second year of each data release cycle of Continuous NHANES. In a recent study of cohort obesity patterns that used NHANES data and the same procedure for calculating birth years, results were robust to alternative specifications of *period* (J. M. Lee et al. 2010). *Age* is the age of the individual, in completed years, at the time of the survey. To ensure large enough age-cohort cells, we analyze cohorts born in ten-year-wide intervals (1910 to 1919, 1920 to 1929, etc.). Using this approach, we obtain a total of 8 ten-year birth cohorts between 1910-1919 and 1980-1989. This method involves assuming that upon reaching age 20, diabetes prevalence is not affected by migration. We test the sensitivity of our results to this assumption by excluding foreign-born individuals from the sample.

***Statistical methods***

Prevalence is calculated as the proportion of individuals in the given age-period or age-cohort cell with diabetes as defined above. Calculations are adjusted for complex survey design using strata and primary sampling units provided by the National Center for Health Statistics (NCHS), along with survey weights. For HbA1c, we use the final examination weight provided by NCHS; because we pool adjacent data release cycles of Continuous NHANES, we divide the examination weights in Continuous NHANES by 2, as recommended by NCHS (National Health and Nutrition Examination Survey 2006).

We use least squares to model the age-, cohort- and period-patterns of diabetes prevalence in the U.S. population, using the following regression models:

*Age/Cohort model:* $$ln(Y_{ia}) = \alpha + \beta_a X_a + \beta_i X_i \quad [1]$$

In this equation, $Y_{ia}$ is the proportion of the population in cohort *i* at age *a* with diabetes, $X_a$ is a dummy variable indicating that the observation pertains to age *a*, and $X_i$ is a dummy variable indicating that the observation pertains to cohort *i*.

*Age/Period model:* $$ln(Y_{ia}) = \alpha + \beta_a X_a + \beta_p X_p \quad [2]$$

Here, $Y_{ia}$ and $X_a$ are defined as in Equation 1 and $X_p$ is a dummy variable indicating that the observation pertains to period *p*.

*Age/Period/Cohort model:* $$ln(Y_{ia}) = \alpha + \beta_a X_a + \beta_p X_p + \gamma Coh\_ob \quad [3]$$

$Y_{ia}$, $X_a$, and $X_p$ are defined as in Equation 2 and *Coh_ob* is a continuous variable representing the prevalence of obesity at age 25 in the cohort corresponding to the given age and period. We use age 25 because NHANES inquired about weight at that specific age. Obesity at age 25 serves as a measure of a cohort's history of obesity. The use of a continuous variable to represent birth cohort influences avoids the identification problem that any two of age, cohort, and period indicators can be linearly combined to produce the third (Mason & Fienberg 1985).

Each prevalence estimate is weighted by the number of observations that gave rise to it in order to give more weight to values estimated with greater precision. The log specification implies that cohort membership affects an age-pattern of prevalence multiplicatively; a simple additive specification would suggest implausibly that the age-pattern of prevalence is shifted up or down by the same amount at each age, with no cohort effect on prevalence beyond the earliest age.

Birth-cohort obesity prevalence is estimated using age-25 weight and height recall data in Continuous NHANES waves 1999-2008. Height recall was only asked of

participants aged 50 and over; for younger individuals we used self-reported current height. We identify birth cohorts by subtracting age from survey year, using the beginning of the second year of each of the waves (e.g., 2000.0 for 1999-2000) and aggregate them into five-year wide intervals. The earliest and most recent birth cohorts for whom cohort obesity is calculated are the 1920-1924 and 1975-1979 birth cohorts, respectively. Thus, the age-period-cohort model excludes prevalence estimates that drew exclusively from the oldest or youngest birth cohorts (born 1910-1919 and 1980-1989). **Appendix A.1.1** shows a table of the cohort-obesity prevalence values used in this study.

The examination of diabetes prevalence within birth cohorts allows us to estimate the age-specific incidence of diabetes. In essence, this estimate is made by dividing the prevalence of non-diabetes in a birth cohort at one age interval (e.g. 50 to 54) by the prevalence of non-diabetes in the same birth cohort in the adjacent, younger age interval (e.g. 45 to 49) and adjusting for the fact that people without diabetes die at lower rates than the general population. The prevalence estimates used in this calculation are based upon the age coefficients estimated from the age/cohort model, presented in Figure 3B. These summarize the age-pattern of prevalence revealed within eight birth cohorts, adjusting for cohort-specific effects. Life tables for individuals without diabetes and for the general population are estimated using pooled data from NHANES III and Continuous NHANES (1999-2004 waves) cohorts linked to deaths in the National Death Index through 2006 (National Health Interview Survey (1986-2004) 2009). A discrete hazards model on a person-month file is employed to generate the underlying risks for

predicting mortality rates. The model is implemented on baseline ages 20-74. There were 2,903 deaths among 25,971 respondents.

***Derivation of formula for diabetes incidence in a cohort***

Suppose that 20% of a cohort has diabetes at age 30 and 25% of that cohort has diabetes at age 35. Then the incidence of diabetes (number of new cases per diabetes-free member of the population) between ages 30 and 35 is approximately .05/.80=.0625. That figure refers to incidence over a five-year period, whereas incidence is normally measured annually. An annualized rate would be .0625/5=.0125. This figure is based on the number who are free of diabetes at the beginning of the interval, whereas incidence is typically measured using a denominator measured at the middle of the interval. So the corrected figure is (.05/.775)/5=.0129.

This calculation makes three basic assumptions: (1) Migration does not affect birth cohort prevalence; (2) Those with diabetes at age 30 do not become diabetes-free by age 35, and (3) Those with diabetes at age 30 have the same probability of dying by age 35 as those who were diabetes-free at age 30. In constructing our estimates of the incidence of diabetes, we retain assumption 1 and 2, that migration does not affect prevalence and that those who enter the diabetic state leave it only by death (see Discussion section for analysis of remission). To check the sensitivity of our results to Assumption 1, we excluded foreign-born individuals from our sample, and results were not substantially altered **(see Appendix A.1.6).** We address Assumption 2 in the Discussion section below. However, Assumption 3 is demonstrably untenable (Stokes &

Mehta 2013). Accordingly, our estimates of diabetes incidence adjust for the higher mortality of those with diabetes.

We develop the estimation formula first by referring to the population at exact ages and then substituting equivalent formulas for the population at discrete age intervals. Under our assumptions, the diabetes-free population is subject to two sources of decrement, incident diabetes and death (Preston et al. 2001).

$$ {}_5p_x^0 = \exp\left(-5(\mu_x^0 + \delta_x^0)\right), \qquad [4] $$

where

${}_5p_x^O$ = probability of surviving in the disease-free state from age x to age x+5 for a person free of diabetes at age x,

$\mu_x^O$ = death rate at age x for a person free of diabetes,

$\delta_x^O$ = rate of acquiring diabetes (incidence rate) at age x for a person free of diabetes.

$$ {}_5p_x = \exp(-5\mu_x), \qquad [5] $$

where

${}_5p_x$ = probability of surviving from age x to age x+5 for a randomly-chosen member of the population

$\mu_x$ = death rate at age x for a randomly-chosen member of the population.

Equations 4 and 5 assume that death rates and the incidence rate of diabetes are constant in the age interval x to x+5, producing the exponential functional form.

Call the non-diabetes population at age x $N_x^O$ and the total population at age x $N_x$. Then the prevalence of non-diabetes at age x is

$$\Pi_x = \frac{N_x^O}{N_x}.$$

The prevalence of non-diabetes in the same cohort at age x+5 is

$$\Pi_{x+5} = \frac{N_{x+5}^O}{N_{x+5}} = \frac{N_x^O \exp[-5(\mu_x^O + \delta_x^O)]}{N_x \exp[-5\mu_x]}$$

$$= \Pi_x \exp(-5\delta_x^0) \exp\big(-5(\mu_x^0 - \mu_x)\big). \qquad [6]$$

Rewriting equation 6 gives

$$\exp(-5\delta_x^0) = \frac{\Pi_{x+5}\exp(-5\mu_x)}{\Pi_x \exp(-5\delta_x^0)}$$

or

$$\delta_x^0 = -\frac{1}{5}\ln\left(\frac{\Pi_{x+5} * {}_5p_x}{\Pi_x * {}_5^M p_x^0}\right), \qquad [7]$$

where

${}_5^M p_x^O =$ probability of surviving the risk of death from x to x+5 for a diabetes-free person at age x.

Equation 7 shows that the incidence rate of diabetes between ages x and x+5 can be derived from the ratio of non-diabetes prevalence at x and x+5 and from differences in the survival probabilities between the entire population and the diabetes-free population over that age span. It also shows why a moving average of incidence estimates made using this equation is appropriate: errors in prevalence estimates at any particular age will

appear in the numerator of one age-specific incidence estimate and in the denominator of the adjacent incidence estimate.

Substituting expressions for discrete five-year intervals into the equivalent terms in **Equation 7** gives

$$ {}_{10}\bar{\delta}^{O}_{x} = -\frac{1}{5}\ln\left[\frac{{}_{5}\Pi_{x+5}}{{}_{5}\Pi_{x}}\frac{{}_{5}L_{x+5}\,/\,{}_{5}L_{x}}{{}^{M}_{5}L^{O}_{x+5}\,/\,{}^{M}_{5}L^{O}_{x}}\right], \qquad [8] $$

where

${}_{10}\bar{\delta}^{O}_{x}$ = rate of developing diabetes for a non-diabetic person in the age interval x to x+10,

${}_{5}\Pi_{x}$ = prevalence of non-diabetes at ages x to x+5

${}_{5}L_{x}$ = person-years lived between ages x and x+5 in a life table for the population

${}^{M}_{5}L^{O}_{x}$ = person-years lived between ages x and x+5 in a life table for persons free of diabetes.

We interpret ${}_{10}\bar{\delta}^{O}_{x}$ as pertaining to the age interval x+2.5 to x+7.5, i.e. the five-year age span at the middle of the ten-year age interval x to x+10. We use equation 8 for our incidence estimates in this study, assuming the incidence rate and differential mortality are constant within the five-year age intervals used. Values of ${}_{5}\Pi_{x}$ are calculated from fitted values in the age-cohort model of prevalence. Values of ${}_{5}L_{x}$ and ${}^{M}_{5}L^{O}_{x}$ come from the life tables as described above in the Statistical Methods section.

All statistical analysis was performed using Stata version 11 (StataCorp, College Station, TX). Standard errors were estimated using first-order Taylor series linearization.

## 1.3 RESULTS

### *Prevalence Estimates and Modeled Age and Cohort Patterns*

**Figure 1-1** plots estimates of age-specific diabetes prevalence during the four observation periods under study. The underlying values and their standard errors are reported in **Appendix Tables A.1.2a and A.1.2b**. As reported elsewhere (Y. J. Cheng et al. 2013), there is a general upward trend in prevalence at each age.

**Figure 1-1** shows a pattern in which the prevalence of diabetes declines at some set of ages above 60-64 in each of the four periods. Such a decline could be produced by higher mortality rates among those with diabetes than among those without. However, we show below that this pattern of decline with age is not present when prevalence rates are arrayed by birth cohort. In other words, the declines in prevalence with age in Figure 1-1 result from the increasing prevalence of diabetes among later-born cohorts.

**Figure 1-2** presents estimates of diabetes prevalence among birth cohorts. It is clear that prevalence is rising from one birth cohort to the next, even at younger ages where prevalence is low. Furthermore, prevalence continues to rise even at the oldest ages, which is consistent with a continued positive incidence of diabetes as cohorts age. Declining prevalence with age, a pattern suggested by period data, is not observed among real birth cohorts as they age.

The age-pattern of diabetes, as well as changes in diabetes prevalence from birth cohort to birth cohort, are summarized by our statistical model. **Figure 1-3** plots the coefficients for each birth cohort in the age/cohort regression model. That the coefficients are monotonically increasing shows that more recent birth cohorts have higher diabetes

prevalence than older cohorts. The increase is exceptionally rapid among cohorts born after 1950-59. The implication of the cohort coefficients is that the prevalence of diabetes at any age for the cohort born 1980-89 will be nearly triple that of the cohort born in 1950-59 and 4.9 times that of the cohort born in 1910-19 (derived from **Appendix Table A.1.3a**).

Just as the age/cohort model produces rapidly increasing cohort effects, the age/period model produces rapidly rising period effects. This nearly straight-line increase in prevalence across periods is shown in **Figure 1-4** (see **Appendix Table A.1.3b** for actual values). By themselves, there is nothing in Figures 1-3 and 1-4 that would indicate which model is preferred. Both models produce $R^2$ values above 0.94. But when we add a cohort variable to the age/period model, the prevalence of obesity at age 25, the period effects nearly disappear, as shown in Figure 1-4 (**Appendix Table 3c**). They also become statistically insignificant.

**Figure 1-5** compares the age-patterns of diabetes prevalence that are produced by the age/cohort model, the age/period model, and the age/period/cohort model. By far the most level age pattern is produced by the age/period model. As argued earlier, that age pattern is misleading because it fails to account for the rise in diabetes prevalence from one birth cohort to the next. As was suggested by a comparison of Figures 1-1 and 1-2, the age pattern of diabetes prevalence in a birth cohort is steeper than that in a period. The age-pattern in the age/period model becomes much steeper when birth-cohort obesity is introduced, as shown in Figure 1-5. The age-pattern identified in the age/period/cohort

model is very similar to that in the age/cohort model. The same pattern of results holds when FPG rather than HbA1c is used to define diabetes **(Appendix A.1.4)**.

***Incidence estimates***

Based on Equation 8 above, **Figure 1-6** shows the age pattern of diabetes incidence that is implied by the age pattern of prevalence that we have uncovered. The values on the graph apply to the cohort born 1950-1959, but the shape of the curve is the same for all birth cohorts. The age-pattern of incidence rises to a peak in the age interval 55 to 64 (centered at age 60) and then declines slowly. At its peak from ages 55 to 64, for the cohort born 1950-1959, approximately 1.1% of the diabetes-free population will develop diabetes each year. **Appendix A.1.5** presents numerical details of our incidence estimates. As shown in **Appendix A.1.6,** the age pattern of diabetes incidence is similar when foreign-born subjects are removed from the sample, suggesting that our results are not sensitive to the assumption that migrants experience the same relevant rates as native-born individuals.

***Sensitivity of results to threshold choice***

To examine the sensitivity of results to the choice of the HbA1c threshold, we adopt a threshold of HbA1c levels $\geq 6.0\%$. Recent guidelines from the American Diabetes Association consider individuals at this level to be at "very high risk" of incident diabetes (American Diabetes Association 2012). See **Appendix A.1.7** for a discussion of this choice of threshold. Using this lower threshold, we estimate the

prevalence of being "at least at high risk" of diabetes over time and across birth cohorts, as shown in **Appendix Figures A.1.7a and A.1.7b**. 7,370 individuals in our sample met the more inclusive criterion. A comparison of Figure 1-2 to Appendix Figure A.1.7b shows that the increase across birth cohorts in age-specific prevalence of "at least high-risk" is even more striking than that using the higher cut-off. In particular, the higher prevalence seen in more recent birth cohorts appears at earlier ages in "at least high-risk" than it does in diabetes itself.

We also estimate age/period, age/cohort, and age/period/cohort models of "at least high-risk" prevalence. The patterns described above are largely replicated using the lower cut-off. Consistent with the higher level of prevalence, the rise in prevalence across ages and birth cohorts is greater when HbA1c $\geq$ 6.0% is used. However the introduction of obesity at age 25 into the age/period model has much the same effect as when HbA1c $\geq$ 6.5% is used; it steepens the age effects and reduces the period effects, though a significant period effect remains in the most recent period (see **Appendix Figures A.1.7c-7e** and Figure 1-4). Once again, this result places the spotlight on birth cohort influences in the rise of diabetes in the United States. **Appendix Tables A.1.7a-7c** present numerical details of the results of our modeling of the prevalence of HbA1c $\geq$ 6.0%.

## 1.4 DISCUSSION

Birth cohorts are an attractive vehicle for investigating changes in the prevalence of diabetes because prevalence at any age is a cumulative product of influences in the

past. These influences manifest themselves over the lifetime of birth cohorts, creating close associations in the prevalence of diabetes across age within a cohort.

We show that the prevalence of diabetes in the United States is rapidly increasing from one birth cohort to the next. We demonstrate this increase graphically and by means of an age/cohort model. The increase is especially rapid across cohorts born after 1950-59.

Our results also reveal that the pattern of increase with age in the prevalence of diabetes is considerably faster within a birth cohort than it is across ages in a particular period. The increase with age during any particular period is too mild, or even negative, because it does not account for the higher levels of diabetes evident among more recent birth cohorts.

An additional suggestion of the importance of birth cohort influences on diabetes prevalence is supplied by our age/period/cohort model. While an age/period model shows sharply increasing period effects, the addition of a term measuring birth cohort obesity at age 25 renders the period effects small and insignificant. This result indicates that birth cohort influences – in particular, birth cohort obesity levels – are important determinants of diabetes prevalence.

An innovation of our approach is that we convert estimates of birth cohort diabetes prevalence to estimates of incidence. Such estimates cannot be made using period data alone without the extreme assumption that no population rates are changing (Greenland & Rothman 2008). This assumption is clearly not warranted in the case of diabetes, as shown in Figure 1-1. But such calculations of incidence can be made by

comparing prevalence at different ages for the same birth cohort since any changes in prevalence within a birth cohort must be attributed to some combination of new diagnoses (incidence), differential mortality by diabetes status, and recovery (if any). To estimate incidence, we use the age effect coefficients from the age/cohort model, which is based on observations across eight birth cohorts. We demonstrate that the incidence of diabetes among diabetes-free persons rises steadily to a peak at ages 55 to 64 and then declines slowly.

To the best of our knowledge, these are the first estimates of the age-pattern of diabetes incidence that are based on measured data in a nationally-representative sample. Other estimates of age-patterns of diabetes incidence are few and inconsistent. Age patterns of diabetes incidence that peak and then decline are found in some populations (Berger, Stenström, & Sundkvist 1999; Khan et al. 2011; McDermott, Li, & Campbell 2010; Pavkov et al. 2007). Other studies find that incidence continues to rise with age (Thunander et al. 2008; Wilson, Anderson, & Kannel 1986) or levels off at older ages (Geiss et al. 2006; Monesi et al. 2012). Annual estimates of incidence in the U.S. from the Centers for Disease Control and Prevention (CDC), which are based on retrospective self-reports, show a peak in the age interval 45-64 in some years and at ages 65-79 in other years (Centers for Disease Control and Prevention 2012). Experimental evidence suggests a biological mechanism for increasing incidence with age at the individual level (H.-Y. Lee et al. 2010). One possible explanation for the peak and decline in diabetes incidence in a birth cohort is population heterogeneity in vulnerability to diabetes, with

the most vulnerable individuals being successively selected out of the diabetes-free population as birth cohorts age.

Our study has several limitations. We assume that migration does not affect the prevalence of diabetes in birth cohorts. When we removed foreign-born respondents from the sample, however, the pattern of our results was essentially unchanged (e.g. **Appendix A.1.6**). We also assume no age-cohort interactions. We tested this assumption by including interactions between a continuous variable for age and indicators for the three birth cohorts that provided the most prevalence estimates; coefficients on these interaction terms were not statistically significant ($p>.15$ in all cases).

The small sample sizes in NHANES required us to use ten-year wide birth cohorts and assume homogeneity within those birth cohorts. As a specification check, we divided the birth cohorts into different ten year intervals than reported in this paper (1915 to 1924, 1925 to 1934, etc.). Resulting patterns of prevalence were similar to the results presented here **(**e.g. **Appendix A.1.8).**

The NHANES data do not permit distinguishing between type 1 and type 2 diabetes. However, because type 2 accounts for about 90-95% of all diabetes cases (American Diabetes Association 2012), this was not a serious limitation.
We categorized as diabetic individuals below the 6.5% HbA1c threshold who reported taking medication for diabetes. On the other hand, we did not categorize as diabetic individuals below the 6.5% threshold with self-reported diabetes because we assume that the large majority of this group was assessed using alternative diagnostic criteria, such as FPG or Oral Glucose Tolerance Test. Prior research indicates that relative to these

measures, the HbA1c test identifies as diabetic a smaller group of high-risk individuals (Cowie et al. 2010). For this reason, we did not assume that individuals with self-reported diabetes were ever above the HbA1c threshold for diabetes.

Our method for estimating diabetes incidence assumes that mortality differences between people with and without diabetes have been constant; the literature on this question is unresolved (Gregg et al. 2012; Gregg, Gu, Cheng, Narayan, & Cowie 2007; Stokes & Mehta 2013). Our calculations also assume that there is no remission once the diabetes-defining threshold is reached (i.e. one can only exit the diabetic state through death). Remissions would offset new cases and produce an underestimate of the incidence rate. The principal source of remission of diabetes is bariatric surgery. According to the American Society for Metabolic and Bariatric Surgery (ASMBS), the number of procedures reached 103,000 in 2003 (National Institute of Diabetes and Digestive and Kidney Diseases 2011). There were approximately 21,708,000 Americans aged 20+ with HbA1c values of 6.5% or greater in that year (Y. J. Cheng et al. 2013; U.S. Census Bureau 2005). Assuming that all those who had the surgery had diabetes, the annual rate of surgery among people with diabetes was .00497 in 2003. Two recent randomized clinical trials investigated the efficacy of bariatric surgery among those with diabetes. One found a one-year success rate in reducing HbA1c below 6.0% of 42% (Schauer et al. 2012) and the other a two-year rate of success of reducing HbA1c below 6.5% of 75% (Mingrone et al. 2012). If we assume that the higher figure applies to the 5-year success rate required in our calculations, bariatric surgery would produce a remission rate of (.75)(.00497) = .00373 among people with diabetes in 2003. Since the

ratio of people without diabetes to people with diabetes in that year was 9.12 (Y. J. Cheng et al. 2013), the rate of flow into the non-diabetic population as a result of successful bariatric surgery was .00373/9.12 = .00041. This value compares to an incidence rate above age 50 of about .010 in our calculations. So the incidence rate above age 50 would be perhaps higher by the factor 1.04 if allowance were taken of remission from bariatric surgery.

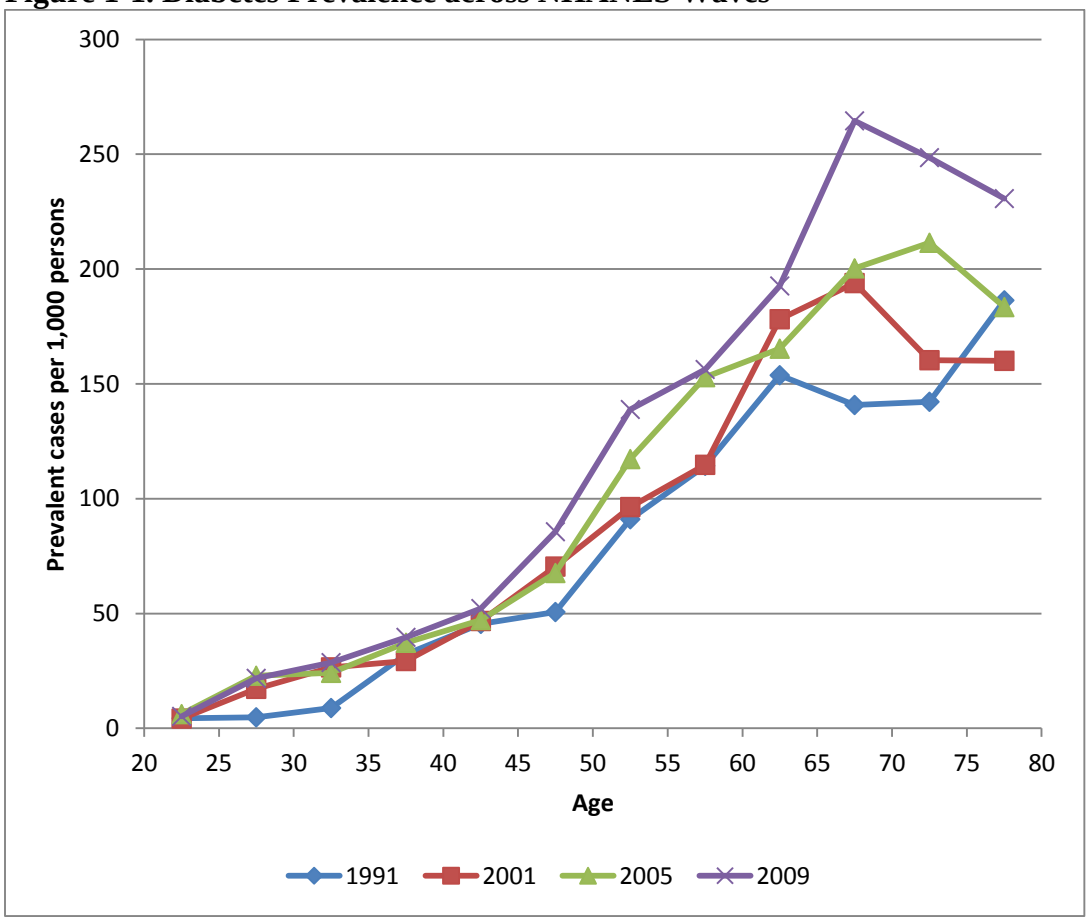
There are other sources of remission, of course, but in these two randomized clinical trials the remission rates for very intensive non-surgical medical treatment was only 12% (Schauer et al. 2012) and 0% (Mingrone et al. 2012). Due to the intensive nature of the medical treatment, these findings can be considered an upper bound on remission rates in the diabetic population at large. It is worth noting that projections of future diabetes prevalence assume the cure rate for diabetes is zero (Boyle, Thompson, Gregg, Barker, & Williamson 2010), and clinical guidelines imply that people who have been diagnosed with diabetes are considered diabetic even if their blood glucose is under control (Ali et al. 2013).

Two recent studies of individuals in NHANES found that secular changes in time-of-survey BMI explained some but not all of the secular increase in the prevalence of diabetes and prediabetes (Bullard et al. 2013; Y. J. Cheng et al. 2013). Our findings also implicate the rise in obesity for increases in diabetes but we use aggregate data on birth cohorts and an historical rather than contemporary indicator of obesity. That both current and past levels of obesity affect an individual's risk of developing diabetes has been demonstrated in prior research (Abdullah et al. 2011). Thus, our results are consistent

with other analyses that identify increases in the prevalence of obesity as an important factor in the rise in diabetes.

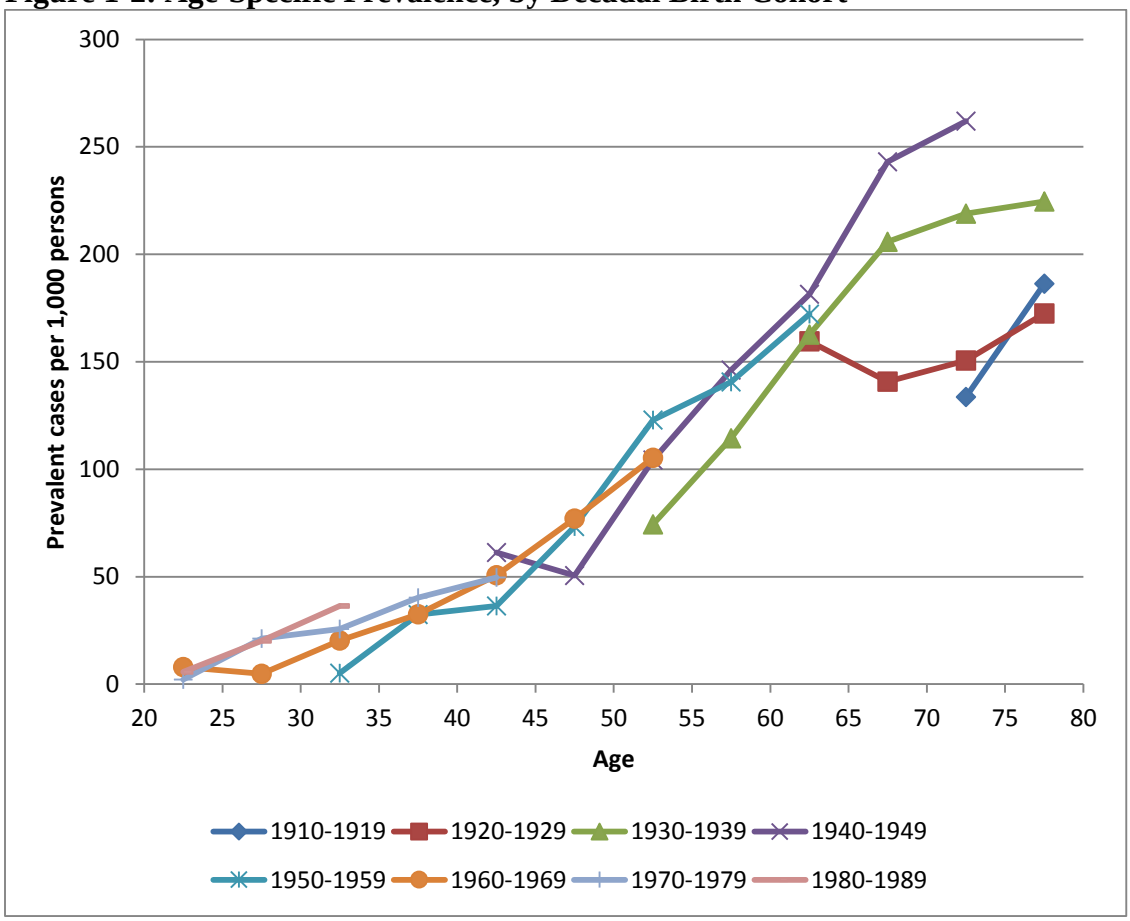
The prevalence of obesity has increased dramatically across recent US birth cohorts. We have shown that birth-cohort prevalence of diabetes is associated with birth-cohort levels of obesity at age 25. Because cohort effects persist as cohorts age, our results suggest that diabetes prevalence is likely to continue increasing despite an apparent plateauing of obesity in recent years (Flegal et al. 2012). Additional analyses should investigate the implications of the birth cohort trends identified here for future diabetes prevalence in the United States.

**Figure 1-1. Diabetes Prevalence across NHANES Waves**



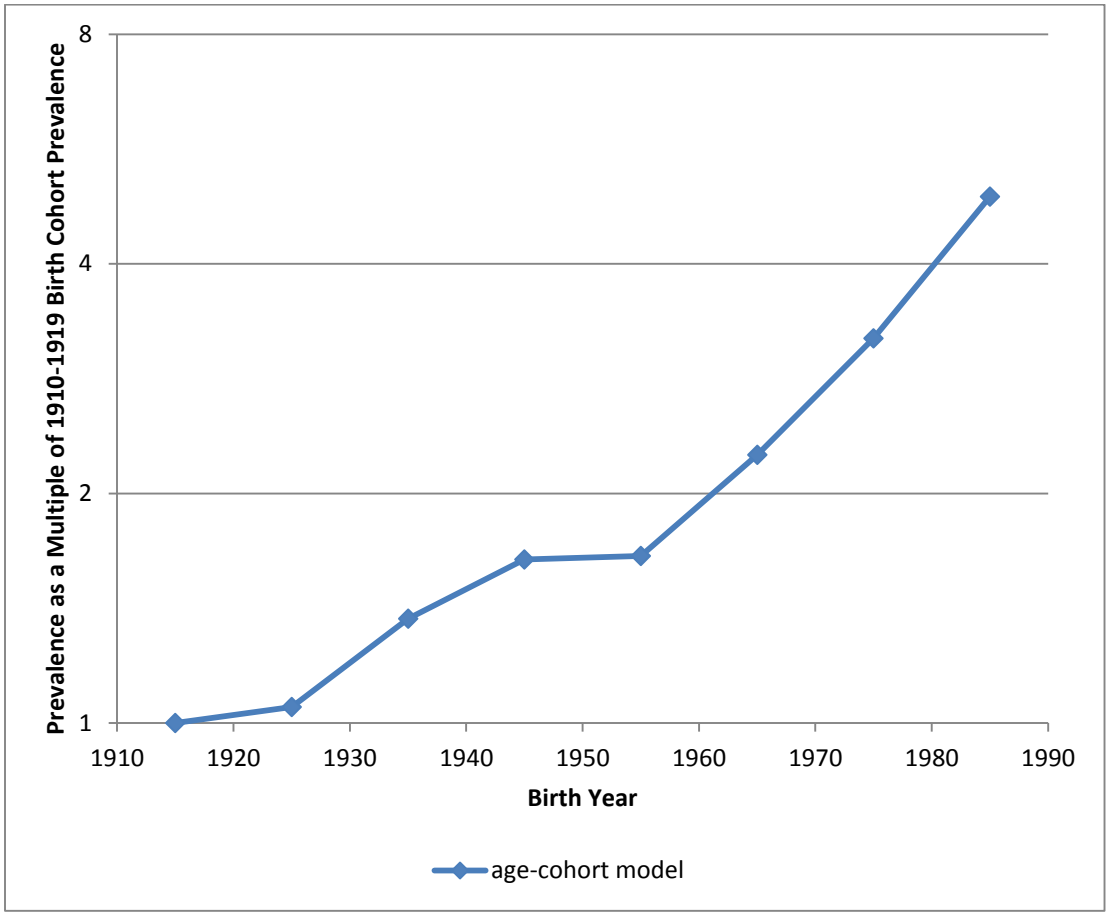

Data: National Health and Nutrition Examination Surveys (NHANES), United States, 1988-1994 and 1998-2010. N=40,130. Diabetes defined as HbA1c at least 6.5% or taking diabetes medication. Values are weighted to reflect the contemporary U.S. population.

**Figure 1-2: Age-Specific Prevalence, by Decadal Birth Cohort**



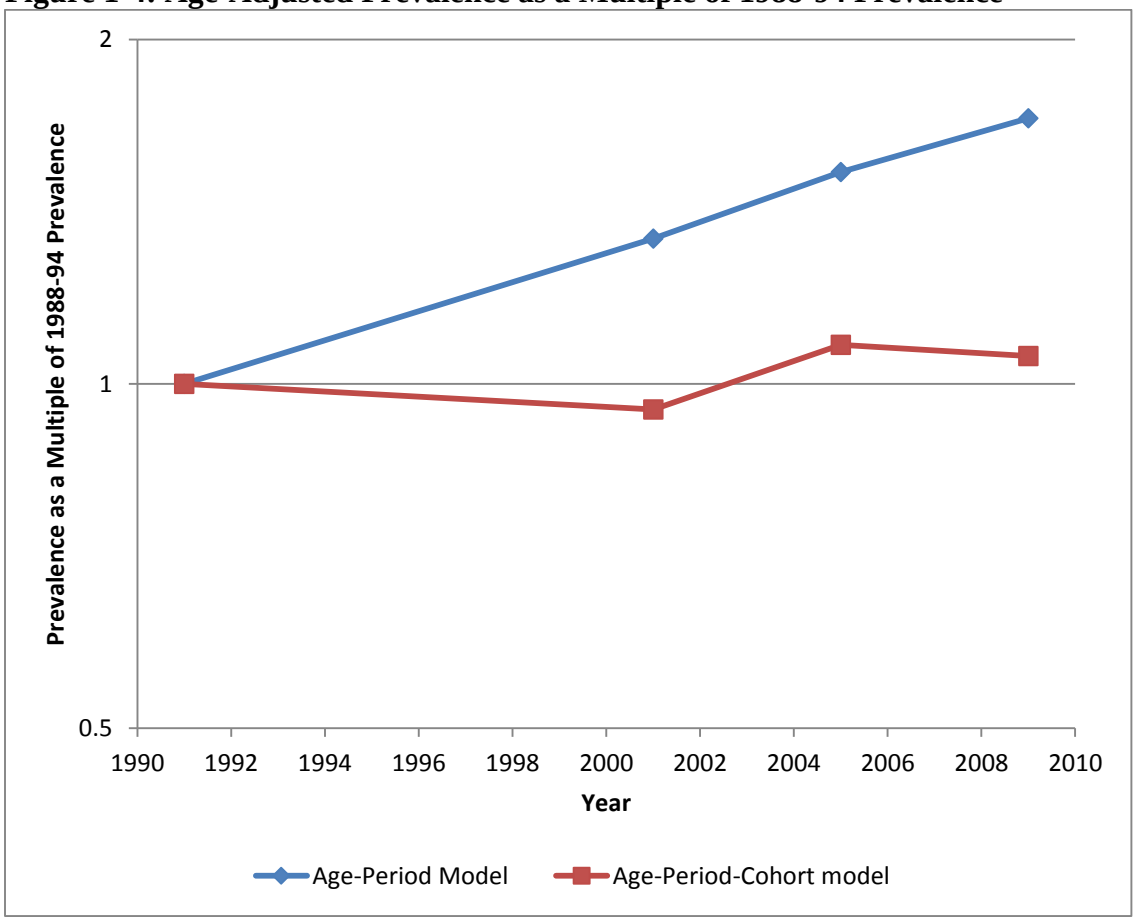

Data: National Health and Nutrition Examination Surveys (NHANES), United States, 1988-1994 and 1998-2010. N=40,130. Diabetes defined as HbA1c at least 6.5% or taking diabetes medication. Values are weighted to reflect the contemporary U.S. population.

**Figure 1-3: Age-Adjusted Diabetes Prevalence in Birth Cohorts Relative to those born 1910-19**



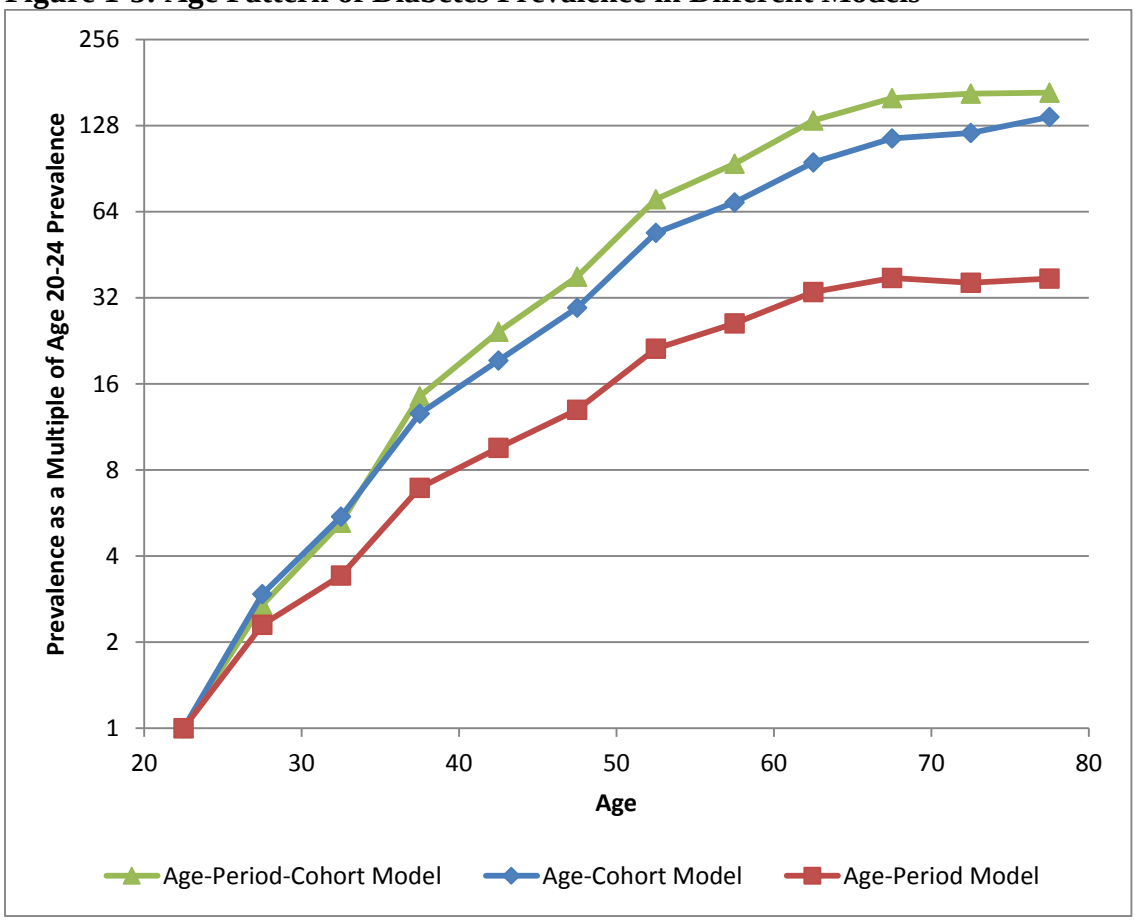

The graph shows the age-adjusted prevalence of diabetes in each birth cohort as a multiple of the age-adjusted prevalence for the 1910-1919 birth cohort.
Data: National Health and Nutrition Examination Surveys (NHANES), United States, 1988-1994 and 1998-2010. N=40,130. Diabetes defined as HbA1c at least 6.5% or taking diabetes medication. Estimates are weighted to reflect the contemporary U.S. population.

**Figure 1-4: Age-Adjusted Prevalence as a Multiple of 1988-94 Prevalence**



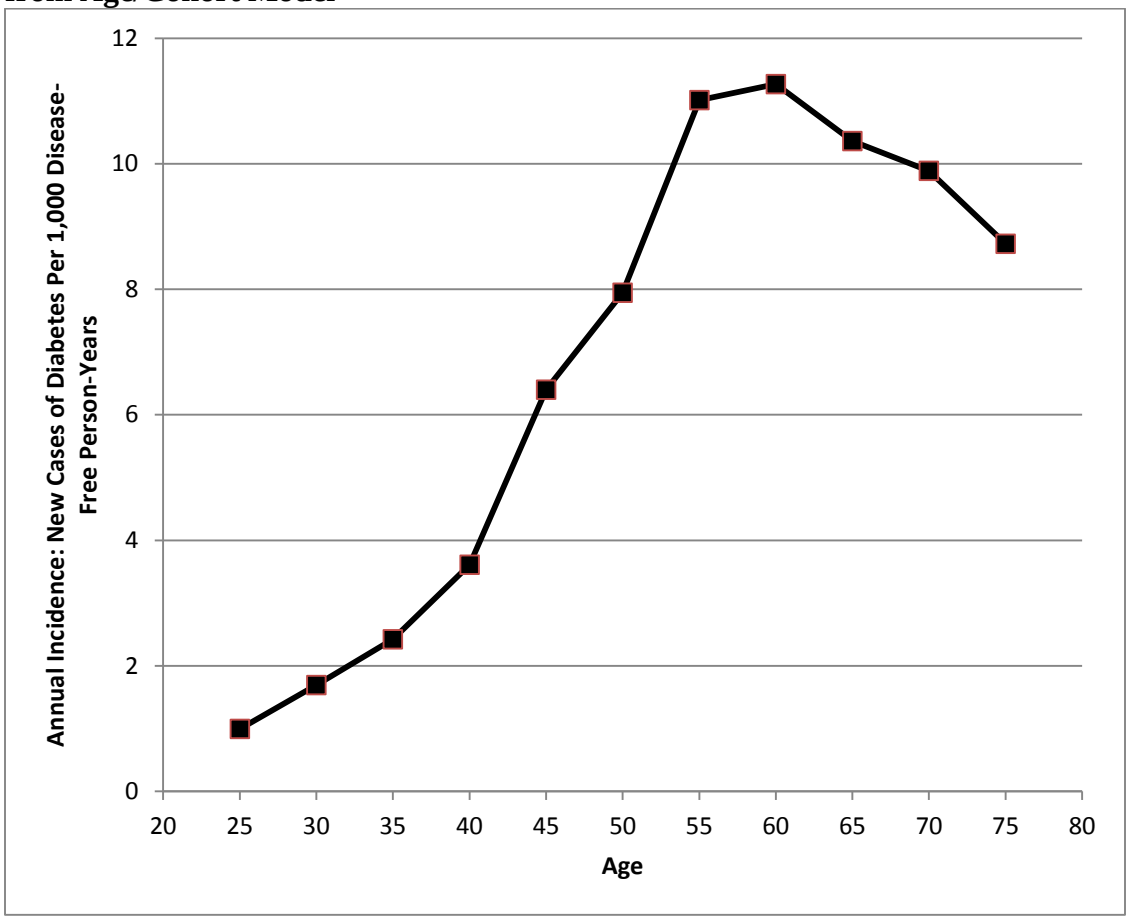

This figure shows age-adjusted prevalence as a multiple of age-adjusted prevalence in the reference 1988-1994 period for the Age/Period model (diamonds) and Age/Period/Cohort model (squares). Data: National Health and Nutrition Examination Surveys (NHANES), United States, 1988-1994 and 1998-2010. N=40,130. Diabetes defined as HbA1c at least 6.5% or taking diabetes medication. Values are weighted to reflect the contemporary U.S. population.

**Figure 1-5: Age Pattern of Diabetes Prevalence in Different Models**




This figure shows age-specific prevalence as a multiple of the prevalence at age 20-24 in the Age/Cohort model (triangles), Age/Period model (diamonds), and Age/Period/Cohort model (squares). Data: National Health and Nutrition Examination Surveys (NHANES), United States, 1988-1994 and 1998-2010. N=40,130. Diabetes defined as HbA1c at least 6.5% or taking diabetes medication. Values are weighted to reflect the contemporary U.S. population.

**Figure 1-6: Smoothed Age Pattern of Diabetes Incidence Using Prevalence Values from Age/Cohort Model**




Incidence estimated from Age/Cohort model of diabetes prevalence and differential mortality estimates, detailed in Statistical Methods section. Three-term moving average of incidence is plotted. For graphical purposes, incidence values are plotted for the cohort born 1950-1959, but the shape of the curve is the same for all birth cohorts. Data: National Health and Nutrition Examination Surveys (NHANES), United States, 1988-1994 and 1998-2010. N=40,130. Diabetes defined as HbA1c at least 6.5% or taking diabetes medication. Values are weighted to reflect the contemporary U.S. population.

# 2 Lifetime risk of dementia in the United States

## 2.1 INTRODUCTION

Dementia is increasingly recognized as a major source of disease burden in the United States (Murray 2013). A national study estimated that 3.4 million American adults over 70 had dementia, corresponding to a prevalence of approximately 13.9% (Plassman et al. 2007). Dementia imposed a financial cost of over $40,000 per affected person per year, comparable to the financial costs of heart disease and cancer (Hurd et al. 2013). Americans over 60 reported fearing dementia more than any other disease, including cancer (Alzheimer's Association 2014). As the U.S. population ages, the number of Americans with dementia is very likely to increase in the coming decades (He & Larsen 2014; Kasper et al. 2015). Using nationally representative, longitudinal data, this study will estimate the probability that an average dementia-free person will develop dementia in the course of life.

An approach that incorporates the competing risks of death and dementia incidence in a prospective cohort allows one to estimate the risk that the average dementia-free individual will develop dementia in the future. It can also provide an estimate of the related measures of dementia-free life expectancy and life expectancy with dementia. These quantities are important for individuals, businesses, and governments as they plan for retirement, save and contribute to pensions, and assess future health care costs and caregiving needs. For demographers and epidemiologists, these quantities provide meaningful insight into the question of whether long-run gains in

survivorship are being experienced in healthy or unhealthy states (Crimmins & Beltrán-Sánchez 2010; Crimmins et al. 2009).

Data from the Framingham Heart Study and from a national Canadian sample have been used to report the lifetime risk of developing dementia using a competing-risks framework (Carone, Asgharian, & Jewell 2014; Seshadri & Wolf 2007; Seshadri et al. 1997). Dementia-free life expectancy was reported for a large cohort in the Pacific Northwest, known as the Adult Changes in Thought study (ACT) (Tom et al. 2015). Though informative, the Framingham, Canadian, and ACT cohorts are not representative of the U.S. population. For example, these cohorts had a larger proportion of subjects who were white than did the U.S. as a whole. Studies have generally found that African Americans have higher age-specific rates of mild cognitive impairment and of dementia than whites (Katz et al. 2013; Sheffield and Peek 2011). As the nation gets more racially and ethnically diverse, these cohorts are decreasingly representative of the U.S. elderly population. Therefore, there is a need for estimates of lifetime risk of dementia from nationally representative data.

## 2.2 METHODS

### *Sample and definitions*

This study uses the Aging, Demographics, and Memory Study (ADAMS), a nationally representative, longitudinal study of cognitive health and dementia conducted in four waves from 2001 to 2009 (Langa et al. 2005). ADAMS, a probability subsample of the Health and Retirement Study (HRS), examined adults aged 70 and older with a

series of cognitive, psychological, and neurological tests, and conducted an extensive medical history, an inventory of current prescription medications, a neurology-focused physical exam, and a family/caregiver questionnaire. The testing was conducted in person by trained technicians and nurses and supervised by neuropsychologists (Langa et al. 2005). Diagnostic criteria were based on the Diagnostic and Statistical Manuals of Mental Disorders, DSM-III-R and DSM-IV, and final diagnosis of dementia was made by a consensus expert panel of physicians (Heeringa et al. 2009; Langa et al. 2005). Detailed descriptions of the ADAMS sample and assessment tools have been previously published (Heeringa et al. 2009; Langa et al. 2005; Plassman et al. 2007).

According to the DSM, the essential feature of dementia is the development of multiple cognitive deficits that include memory impairment and at least one of aphasia (language deficit), apraxia (movement deficit), agnosia (deficit in recognition of objects or senses), or executive functioning deficit (American Psychiatric Association 2000). The cognitive deficits must represent a decline from past abilities and must be severe enough to cause impairment in occupational or social functioning (American Psychiatric Association 2000). The most common type of dementia is Alzheimer's disease (AD), which accounts for 60% to 80% of dementia cases. The next most common type is vascular dementia, which alone accounts for about 10% of cases but which is often found together with AD (Alzheimer's Association 2014).

The initial wave of ADAMS, 2001-2003, examined 856 subjects to generate baseline estimates of dementia prevalence in the U.S. (Plassman et al. 2007). The subsequent waves followed 456 dementia-free individuals for dementia incidence

(Plassman et al. 2011). The second wave focused on subjects whose baseline status was "cognitively impaired, no dementia"; this second wave assessed subjects 16 to 18 months after their baseline assessment. For the third and fourth waves, all living subjects who were dementia-free at baseline were in the sampling frame. Subjects in the third wave averaged 3.7 years since their most recent assessment, and subjects in the fourth wave averaged 1.8 years since their most recent assessment (Plassman et al. 2011). Despite the relatively long intervals between assessments, ADAMS investigators could determine, based on informant reports, medical records, and clinical assessment, that a subject experienced the onset of dementia at any time since the previous assessment. For example, if a 72 year old subject was deemed dementia-free at baseline and then assessed at age 76 and found to have dementia, investigators could determine that his age at the onset of dementia was 73. The assignment of ages at dementia onset during the inter-assessment interval allows for the estimation of dementia incidence rates, rather than probabilities. Thus the ADAMS data can be used to calculate age-specific incidence of dementia, an essential ingredient in making estimates of lifetime risk.

Incidence rates in large age categories (below 80, 80 to 89, and 90+) have been published in a prior study (Plassman et al. 2011). In that study, incidence was measured as follows. First, the number of incident cases in the ADAMS sample, by age group, was counted. This number was converted to a comparable figure for the U.S. population using the sampling weights. Then, the number of person-years at risk for each ADAMS subject, by age group, was determined as the number of years from the first ADAMS assessment to the first of the following events: (1) dementia onset, (2) death, or (3) completion of the

final ADAMS assessment. Finally, the number of person-years at risk in the sample was converted to a comparable figure for the population using the sampling weights, and the number of new cases in the population was divided by the number of person-years at risk in the population (Plassman et al. 2011). Mortality rates were not specifically reported in that study.

Mortality data for the current investigation come from ADAMS' link to the Health and Retirement Study's mortality tracking via the National Death Index (NDI), which provides vital status and, if deceased, month of death, as of December 2011. The 856 ADAMS subjects constitute the individuals at risk of mortality. I use the mortality data to generate estimates of the age-specific ratio of mortality rates between those with and those without dementia. Mortality rates for the entire U.S. population come from the Social Security Administration (SSA) cohort life tables (Bell & Miller 2005).

***Demographic methods***

In a stationary population subject to a given life table, any two parameters among disease incidence, prevalence, and differential mortality between those with and those without the disease imply the third parameter (Preston, Heuveline, and Guillot 2001, chapter 4). The ADAMS data allow for the estimation of all three parameters, as discussed above; they therefore enable one to assess the stationary population assumptions of the constancy of age-specific incidence rates and differential mortality over time. Alternatively, if there is a strong basis for assuming the constancy of rates, one can use estimates of two of the parameters along with the assumption of constancy of

rates to derive the third parameter. I will use ADAMS incidence and prevalence estimates and the assumption of stationarity to derive an estimate of differential mortality and to estimate lifetime risk of dementia. Then I will relax the assumption of stationarity and estimate differential mortality directly from the ADAMS data, producing a second set of lifetime-risk estimates.

In the context of ADAMS and dementia, differential mortality is the parameter about which there is least agreement in the literature on age patterns and functional form (Guehne, Riedel-Heller, & Angermeyer 2005). Existing estimates of the ratio of mortality among women age 75 to 84 with dementia to mortality among same-age women without dementia vary from 4.07 in Canada (Ostbye, Hill, & Steenhuis 1999) to 2.59 in Spain (Villarejo et al. 2011). Although this ratio is consistently found to decline with age, the pace of decline varies widely across studies (James et al. 2014; Ostbye et al. 1999; Tschanz, Corcoran, & Skoog 2004; Villarejo et al. 2011), making estimates of the age pattern of differential mortality subject to strong parametric assumptions. The ADAMS team has published estimates of age-specific prevalence and incidence to which I can benchmark my own estimates (Plassman et al. 2007, 2011), but no comparable ADAMS-based estimates of differential mortality exist.

In this study, where ADAMS data go to 2009 and mortality data go to 2011, an additional challenge in estimating differential mortality relates to censoring. Individuals' dementia status is known as of their last ADAMS assessment, but mortality follow-up continues for several additional years, during which time new cases of dementia go unobserved. Thus the question arises of when to censor individuals whose last ADAMS

assessment categorized them as without dementia. If one follows them as long as the mortality follow-up allows, one will misclassify many deaths as deaths without dementia. If one censors these individuals too early, one under-counts deaths and person-years without dementia. In either case, the distortions to the estimated differential mortality are potentially severe. The results will show that estimated differential mortality has the largest standard error among the three parameters, and given underlying questions about functional form and the censoring of individuals not diagnosed with dementia, it is likely that standard errors of estimates of differential mortality do not capture all the uncertainty associated with those estimates.

The difficulties in directly estimating differential mortality motivate the use of the stationarity assumption of constancy of rates, and there is considerable evidence suggesting that age-specific dementia incidence has been constant over the last decades (Asgharian, Wolfson, & Zhang 2006; Ewbank 2004; Rocca et al. 2011). Since my estimates cover only ages 70 and above, the time interval during which I would assume constancy of incidence rates is relatively short. Given the evidence for stationarity and the difficulties directly estimating relative risks of death, I will begin with a method that assumes that age-specific dementia incidence and differential mortality have been constant over time, deriving differential mortality from stationary-population relations rather than estimating it directly.

Other studies do find declines over time in prevalence of moderate or severe cognitive impairment (Langa et al. 2008; Manton & Ukraintseva 2005). Declines in prevalence could be consistent with constant incidence if average duration of dementia

declines, but these contrary findings provide some evidence against the stationarity assumption. I will therefore conduct additional analyses using differential mortality I estimate directly from the ADAMS data – despite the limitations of such estimation – and not assuming a stationary population. Comparing the prevalence estimated at baseline in ADAMS to that implied by my estimated incidence and differential mortality provides an informal test of the stationarity assumptions: if the two prevalence series are concordant, then the incidence and differential mortality that gave rise to baseline prevalence closely aligns with the incidence and differential mortality observed longitudinally.

***Approach 1: Assume constancy of incidence rates over time (stationary-population approach)***

Because of the small sample size in ADAMS, I fit simple models to generate smooth age patterns of dementia prevalence ($P$) and incidence ($h$):

$$logit(P_x) = \alpha + \beta x, \quad [1]$$

$$logit(h_x) = \alpha' + \beta' x, \quad [2]$$

where $x$ is exact age. This model broadly conforms to the functional form of the age pattern of Alzheimer's disease rates, and Alzheimer's prevalence and incidence rates have been shown to have similar functional forms (Brookmeyer & Gray 2000; Brookmeyer et al. 2011; Ziegler-Graham, Brookmeyer, Johnson, & Arrighi 2008). For prevalence, I fit the model using logistic regression on the baseline ADAMS sample (n=856). Baseline age was provided in completed years ("last birthday"), so exact age ($x$) was the reported age plus 0.5.

For incidence, I fit the model using a discrete-time logistic regression on a person-year data file (Allison 1984), using the 456 subjects followed longitudinally. Age of dementia onset was reported in completed years, so, for incident cases, the exact age at incidence was set at the reported age (last-birthday) of onset plus 0.5. Subjects who never received a diagnosis of dementia from ADAMS investigators, including those who died without a dementia diagnosis, were censored. Among the censored subjects, those whose status at the end of the ADAMS study period was "alive, dementia-free" contributed dementia-free person-years up to and including their exact age (in months) at their last assessment. Censored subjects whose status at the end of ADAMS was "died without dementia" contributed dementia-free person years until their exact age at death. For example, if a subject's status at the end of ADAMS was "died without dementia," and she died at age 78 and 5 months, then she contributed person-years of exposure until she was 78.41666. Her death would be assigned to the interval between exact ages 78.0 and 79.0. The approach of carrying the last assessment of deceased individuals forward until death is consistent with previous ADAMS reports (Plassman et al. 2011) and recommendations based on simulations of censored time-to-dementia data (Leffondré, Touraine, Helmer, & Joly 2013). It is based on the idea that if the deceased individuals had survived and developed dementia, the investigators could have been able to observe their dementia onset; decedents were therefore at risk of dementia onset until their deaths. A sensitivity analysis will treat these two forms of censoring – death without dementia and survival without dementia to the end of the study period – in a more consistent

fashion by censoring surviving dementia-free subjects at the end of the ADAMS study period rather than at their last assessment.

There is considerable evidence in the literature suggesting that age-specific incidence rates of dementia do not vary by sex (Chêne et al. 2015; Plassman et al. 2011; Ruitenberg, Ott, & Swieten 2001). When a sex term was included in Equation 2, its coefficient was statistically insignificant (p>0.20).

Call the fitted prevalence vector $({}_nP_x)$, and the incidence vector $({}_nh_x)$; these are the same for males and females. The 1920 birth cohort would have been aged 81 to 88 over the study period of ADAMS, making this cohort's life table a good approximation of the overall level of mortality in the population the ADAMS cohort represents. The SSA life table for this cohort provides mortality rates $({}_nm_x)$, survivors to exact age x $(l_x)$, and person-years lived in each age interval in the entire population $({}_nL_x)$; these values are sex-specific. In other words, the level of mortality varies by sex, which will generate sex-specific estimates of lifetime risk, but the other input quantities are constant across sex. Employing the Sullivan method (Mathers & Robine 1997; Sullivan 1971) and using single-year age intervals, the number of person-years lived without and with dementia, respectively, in the age interval (*x, x+1*) are:

$$ {}_1L_x^{DF} = {}_1L_x(1 - {}_1P_x) \quad [3] $$

$$ {}_1L_x^{D} = {}_1L_x * {}_1P_x. \quad [4] $$

The assumptions of stationarity are sufficient for the Sullivan method to generate unbiased and consistent estimates of person-years lived in each state (Imai & Soneji

2007). Taking $l_{70}$ as the radix for the entire population, I estimate the population with dementia at exact age 70 as

$$l_{70}^{D} = l_{70} * P_{70.0} \quad [5]$$

and dementia-free (*DF*) population as

$$l_{70}^{DF} = l_{70} * (1 - P_{70.0}). \quad [6]$$

I fill the life table as follows, assuming events occur on average halfway through intervals. For survivors:

$$l_{x+1}^{D} = (2 * {}_1L_x^D) - l_x^D, \quad [7]$$

$$l_{x+1}^{DF} = (2 * {}_1L_x^{DF}) - l_x^{DF}. \quad [8]$$

The number of new dementia cases is

$${}_1^{dem}d_x^{DF} = {}_1L_x^{DF} * {}_1h_x. \quad [9]$$

In Approach 1, I assume that individuals do not develop dementia and die in the same single-year age interval. Therefore, the number of deaths from the dementia-free population is

$${}_1^{death}d_x^{DF} = l_x^{DF} - l_{x+1}^{DF} - {}_1^{dem}d_x^{DF}. \quad [10]$$

The death rate among the dementia-free population is

$${}_1^{death}m_x^{DF} = {}_1^{death}d_x^{DF} \Big/ {}_1L_x^{DF}. \quad [11]$$

To derive the death rate among those with dementia, I first decompose the mortality rate in the entire population into a weighted average of the mortality rates of the population with and the population without dementia, where the weights are the prevalence of dementia and its complement:

$$ {}_1m_x = {}_1m_x^D * {}_1P_x + {}^{death}{}_1m_x^{DF} * (1 - {}_1P_x), \quad [12] $$

which can be rearranged as

$$ {}_1m_x^D = \left[ {}_1m_x - {}^{death}{}_1m_x^{DF} * (1 - {}_1P_x) \right] \Big/ {}_1P_x. \quad [13] $$

Then the ratio of mortality rates (with dementia vs. without dementia) as implied by the prevalence, incidence, and stationary population assumption is

$$ {}_1RR_x = \frac{{}_1m_x^D}{{}^{death}{}_1m_x^{DF}}. \quad [14] $$

The primary quantity of interest is the lifetime risk of dementia for an age-*a* person without dementia:

$$ LFTM_a = {\Sigma_{x=a}^{w}}^{dem}{}_1d_x^{DF} \Big/ l_a^{DF}, \quad [15] $$

where *w* is the highest age interval. Also of interest is dementia-free life expectancy, the average number of years a randomly chosen person age *a* can expect to live free of dementia, under current rates:

$$ DFLE_a = {\Sigma_{x=a}^{w}}\ {}_1L_x^{DF} \Big/ l_a . \quad [16] $$

Total life expectancy is as in a single-decrement life table:

$$ LE_a = {\Sigma_{x=a}^{w}}\ {}_1L_x \Big/ l_a , \quad [17] $$

and by construction, unconditional life expectancy with dementia – that is, the average number of years an age-*a* person randomly chosen from the population can expect to live with dementia, under current rates, is

$$ DLE_a = LE_a - DFLE_a. \quad [18] $$

We can also define conditional life expectancy without dementia as the average number of years a *dementia-free* person of a given age can expect to live free of dementia:

$$DFLE'_a = {}^{\Sigma^{w}_{x=a}\ {}_1L^{DF}_x}\Big/_{l^{DF}_a}. \qquad [19]$$

This quantity is valuable because the number of dementia-free person-years lived above age *a* for someone who already has dementia at age *a* is zero, and the people contributing these zeros are counted in the denominator of **Equation 16**; they are not counted in the denominator of **Equation 19**. The conditional DFLE is estimable using this method because all person-years lived in a dementia-free state above age *a* are experienced by people who were dementia-free at age *a*; the numerator and denominator therefore match. In contrast, $(\Sigma^{w}_{x=a}\ {}_1L^{D}_x)/l^{DF}_a$ is *not* equal to the average number of years a dementia-free person age *a* can expect to live with dementia, because some of the person-years lived with dementia above age *a* – years contributing to the numerator – are experienced by people who already had dementia at age *a* and who thus do not contribute to the denominator.

***Approach 2: No assumption of stationarity; estimate differential mortality directly from ADAMS data***

Disease prevalence at a point in time embodies a history of disease incidence and differential mortality. Therefore, age-specific prevalence and incidence estimated in a stationary population – that is, one in which incidence and differential mortality have not changed over time – imply a unique pattern of differential mortality. If the population is non-stationary, then past incidence and differential mortality embodied in current

prevalence estimates do not necessarily convey information about current incidence and differential mortality.

To estimate a current age pattern of differential mortality directly from the ADAMS data, without assuming the population is stationary, I use a Gompertz-type model of death rates as a function of an indicator for dementia presence, exact age *x*, and their interaction, fit with a Poisson regression on a person-year data file (Loomis 2005). Dementia status is modeled as a time-varying indicator to incorporate both baseline prevalent cases and incident cases (Palloni & Thomas 2013). The model is:

$$\ln(m_{x,dem}) = \alpha + \beta_1 x + \beta_2 dementia + \beta_3 x * dementia. \qquad [20]$$

As with the estimation of dementia incidence discussed above, subjects who died without a dementia diagnosis during the ADAMS study period contribute dementia-free person years until their exact age at death, and subjects who survived ADAMS without a dementia diagnosis contributed dementia-free person years until their last ADAMS assessment. Mortality data for the period after ADAMS (2009 to 2011) was used only for those with a dementia diagnosis, whose state could not change until death. Not using mortality data from the post-ADAMS period for individuals without a dementia diagnosis avoids large misclassification errors whereby persons who develop dementia subsequent to ADAMS would wrongly contribute deaths without dementia and person-years without dementia to the calculations.

Based on Equation 20, the ratio of the mortality rate among persons with dementia to that among persons without dementia – also known as the risk ratio, rate ratio, or relative risk (RR) – is

$$RR_x = \exp(\alpha + \beta_1 x + \beta_2 + \beta_3 x) \Big/ \exp(\alpha + \beta_1 x) = \exp(\beta_2 + \beta_3 x). \quad [21]$$

In this way, the *ratio* of the two mortality rates is estimated from the ADAMS sample, but the actual values of the mortality rates can be adjusted to match national data with many more deaths using the SSA 1920 cohort life tables. Consistent with most of the literature, the ratio of mortality rates between those with and those without dementia were held constant across sex (Agüero-Torres, Fratiglioni, & Guo 1999; Garcia-Ptacek et al. 2014; Johnson, Brookmeyer, & Ziegler-Graham 2007; Lönnroos, Kyyrönen, Bell, van der Cammen, & Hartikainen 2013; Meller, Fichter, & Schroppel 1999; Villarejo et al. 2011; Witthaus, Ott, Barendregt, Breteler, & Bonneux 1999). When a sex term and a sex-by-dementia-status interaction term were included in Equation 2, the coefficient on the sex-by-dementia term was not statistically significant ($p>0.30$), providing additional justification for keeping differential mortality constant across sex. As with the modeling of incidence rates, pooling males and females to estimate differential mortality is useful with a small sample size such as in ADAMS. In this model, the only quantity that differed by sex was the overall level of age-specific mortality in the entire U.S. population.

The inclusion of the interaction term with the coefficient $\beta_3$ implies that the excess risk of death associated with having dementia declines (assuming $\beta_3$ is negative) with age (Helmer, Joly, Letenneur, Commenges, & Dartigues 2001). This decline arises from two related but distinct forces. The first is the aging of all the individuals in the cohort: as the underlying risk of death rises with age for everyone, the excess risk of death associated with dementia declines. The second force is heterogeneity in frailty

within each group (Vaupel, Manton, & Stallard 1979). Heterogeneity within population groups selects out the frailest individuals first. This force acts more strongly on the higher-mortality group (people with dementia), leaving heartier individuals remaining. In the context of dementia, heterogeneity could arise from changes with age in the average duration of dementia or changes in the prevalence of the APOE e4 allele in the dementia population relative to that in the non-dementia population (Ewbank 2004). This process is similar to the consistent finding of black-white mortality differentials, which decline at older ages (Eberstein, Nam, & Heyman 2008).

In a sensitivity analysis, subjects who survived ADAMS without a dementia diagnosis were censored at the end of the ADAMS observation period, rather than at their last ADAMS assessment, in parallel with the sensitivity analysis for incidence estimates. By increasing the number of person-years lived without dementia and not changing the number of deaths without dementia, the sensitivity analysis will reduce the estimated mortality rate among the non-dementia population, raising the estimated mortality rate ratio.

For a given age, the mortality rate for the entire population can be decomposed into a weighted average of mortality rates of the diseased and disease-free populations, weighted by the age-specific prevalence of the disease, as in **Equations 12 to 14** above:

$$
\begin{aligned}
{}_1m_x &= {}_1m_x^D * {}_1P_x + {}^{death}_{1}m_x^{DF} * (1 - {}_1P_x) \\
&= {}^{death}_{1}m_x^{DF} * {}_1RR_x * {}_1P_x + {}^{death}_{1}m_x^{DF} * \left(1 - {}_1P_x\right). \qquad [22]
\end{aligned}
$$

The terms can be rearranged to solve for the mortality rate in the dementia-free population:

$$ {}^{death}_{1}m_x^{DF} = {}_1m_x \Big/ ( {}_1P_x * {}_1RR_x + 1 - {}_1P_x ) \quad [23] $$

and in the population with dementia:

$$ {}_1m_x^D = {}^{death}_{1}m_x^{DF} * {}_1RR_x, \quad [24] $$

where the overall mortality rate vector ${}_1m_x$ comes from the SSA life table, the age-specific prevalence is described below (see **Equation 39**), and the mortality rate ratio ($RR_x$) is as above in **Equation 21**.

I then construct a multiple-decrement life table for the population without dementia, incorporating elements of the increment-decrement life table to keep track of a model population with dementia (Preston, Heuveline, & Guillot 2001). As with Approach 1, I use single-year age groups and assume no recovery from dementia. The overall rate of decrement from the dementia-free population is the dementia incidence rate, which comes from **Equation 2**, plus the mortality rate for the dementia-free population:

$$ {}_1m_x^{DF} = {}^{death}_{1}m_x^{DF} + {}_1h_x, \quad [25] $$

and the probability of exiting the dementia-free population at a given age, assuming decrements occur on average halfway through each age interval, is

$$ {}_1q_x^{DF} = {}_1m_x^{DF} \Big/ (1 + 0.5\ {}_1m_x^{DF}). \quad [26] $$

The probabilities of exiting the dementia-free population as a result of dementia onset or death, respectively, are:

$$ {}^{Dem}_{1}q_x^{DF} = {}_1q_x^{DF} * \left( {}_1inci_x \Big/ {}_1m_x^{DF} \right), \quad [27] $$

$$ {}^{Death}{}_{1}q_x^{DF} = {}_{1}q_x^{DF} * \left( {}^{death}{}_{1}m_x^{DF} \Big/ {}_{1}m_x^{DF} \right). \quad [28] $$

Define $l_x^{DF}$ as the number of dementia-free survivors to the xth birthday, so that the number of exits from the dementia-free population, by type of exit, is

$$ {}^{i}{}_{1}d_x^{DF} = l_x^{DF} * {}^{i}{}_{1}q_x^{DF}, \; i = Dem, Death. \quad [29] $$

The number of dementia-free survivors to the next age is

$$ l_{x+1}^{DF} = l_x^{DF} - {}^{Dem}{}_{1}d_x^{DF} - {}^{Death}{}_{1}d_x^{DF}. \quad [30] $$

For an approximation of the prevalence of dementia at age 70, I use the fitted value of prevalence for age 70.0 from **Equation 1** in the life table, obtaining $l_{70}^{DF}$ as in **Equation 6**. After age 70, the population with dementia is tracked as follows. The only way to exit the population with dementia is death, so the probability of death with dementia is

$$ {}_{1}q_x^{D} = {}^{Death}{}_{1}q_x^{D} = {}_{1}m_x^{D} \Big/ (1 + 0.5 \; {}_{1}m_x^{D}). \quad [31] $$

The size of the population reaching the xth birthday with dementia is defined as $l_x^D$, so the number of deaths is

$$ {}^{Death}{}_{1}d_x^{D} = l_x^{D} * {}^{Death}{}_{1}q_x^{D} \quad [32] $$

However, those who develop dementia while age *x* are subject to the risk of death $m_x^D$ once they develop dementia. If they develop dementia halfway through the age interval on average, then the probability of death with dementia for these new cases in that interval is

$$ {}^{Death}{}_{new}q_x^{D} = {}_{1}m_x^{D} \Big/ (2 + 0.5 \; {}_{1}m_x^{D}) \quad [33] $$

and the number of deaths among new dementia cases is

$$ {}^{Death}_{new}d_x^D = {}^{Dem}_{1}d_x^{DF} * {}^{Death}_{new}q_x^D. \quad [34] $$

The size of the population with dementia at the subsequent (exact) age is

$$ l_{x+1}^D = l_x^D + {}^{Dem}_{1}d_x^{DF} - {}^{Death}_{1}d_x^D - {}^{Death}_{new}d_x^D. \quad [35] $$

Person-years lived in the dementia-free state are calculated assuming exits occur linearly within age intervals:

$$ {}_1L_x^{DF} = l_{x+1}^{DF} + 0.5(l_x^{DF} - l_{x+1}^{DF}). \quad [36] $$

Person-years lived in a state of dementia are simply

$$ {}_1L_x^D = {}_1L_x - {}_1L_x^{DF}. \quad [37] $$

Filling in the table for the subsequent age ($x$+1) requires an approximation of the proportion of survivors with dementia in the middle of the age ($x$+1, $x$+2) interval, because the mortality rates in **Equations 22-24** pertain to age intervals rather than exact ages. My approximation again uses the assumption that the number of survivors declines linearly over each one-year interval. I assume that half the attrition recorded from exact ages $x$ to $x$+1 will occur from exact age $x$+1 to the middle of the ($x$+1, $x$+2) interval. I denote approximated number of persons in state $i$ in the middle of the age ($x$+1, $x$+2) interval as ${}_1\hat{L}_{x+1}^i$, while I assume that the $L$ column for the entire population (from SSA) records all survivors in the middle of the given age interval:

$$ \begin{aligned} {}_1\hat{L}_{x+1}^D = {}_1L_{x+1} - {}_1\hat{L}_{x+1}^{DF} &= {}_1L_{x+1} - [l_{x+1}^{DF} - 0.5(l_x^{DF} - l_{x+1}^{DF})] \\ &= {}_1L_{x+1} - [1.5l_{x+1}^{DF} - 0.5l_x^{DF}]. \end{aligned} \quad [38] $$

Prevalence of dementia at the subsequent age is estimated as the proportion of mid-interval survivors living in a state of dementia:

$$ {}_1P_{x+1} = {}^{1\hat{L}^D_{x+1}}\Big/ {}_1L_{x+1}, \quad [39] $$

which is used to solve for the mortality rate in the dementia-free population for the age *x+1* interval, using **Equations 22 and 23**.

Because I use narrow (one-year) age intervals, the resulting $\hat{L}^i_x$ columns from **Equation 38** will be close to the $L^i_x$ columns from **Equations 36-37**. (The similarity of the two columns is shown in **Appendix A.2.4)**. The age schedule of prevalence as calculated in **Equation 39** can be compared to that estimated in baseline ADAMS in **Equation 1** as an informal test of stationarity, under the assumption that the model of differential mortality is correct.

Once the multiple-decrement life table is completed, the summary quantities of interest – lifetime risk of developing dementia, unconditional expectancies, and certain conditional expectancies – can be calculated as in **Equations 15 through 19**.

***Simulated reductions in mortality, and simulated delays or reductions in dementia incidence***

Approach 2, which does not assume the constancy of rates over time, lends itself to simulations of future lifetime risk based on changes in mortality or dementia incidence. To assess lifetime risk for younger, lower-mortality cohorts, a secondary analysis used values for ${}_1m_x$ from the 1940 cohort life tables from SSA. The results estimate lifetime risk using current incidence rates and mortality rate ratios as estimated in ADAMS,

isolating the role of declines in the overall level of mortality in changing lifetime risk of dementia.

I also estimate the lifecycle effects of an intervention that delays the onset of dementia. I consider interventions that vary along two parameters: δ, the length of the delay of dementia onset in years, and π, the proportion of the population at risk for whom the intervention is effective. I model these interventions by splitting the model dementia-free population into groups of size $l_{70}^{DF}(1-\pi)$ and $l_{70}^{DF}(\pi)$, subjecting the first group to the dementia incidence rates as modeled in **Equation 2**, and subjecting the second group to the dementia incidence rates as modeled by

$$logit(h'_x) = \alpha + \beta(x-\delta), \qquad [40]$$

where δ is the number of years of delay of dementia onset induced by the intervention. This equation assigns what had been the age-70 incidence rate to age 70+ δ, what had been the age-71 incidence rate to age 71+ δ, and so forth.

Another type of intervention would reduce the risk of dementia at every age, rather than delaying its onset. Such an intervention generates an incidence equation such as:

$$logit(h''_x) = \alpha + (\beta k)x, \qquad [41]$$

where *k* is a value between 0 and 1 that represents the extent to which dementia incidence rises less steeply with age due to the intervention. The closer *k* is to zero, the more effective is the intervention in the sense of reducing the acceleration of dementia incidence. I simulate an intervention where *k* = 0.9, to reduce the (logit of) acceleration of dementia incidence with age by 10%.

Both the dementia-free and with-dementia populations are subject to the same mortality rates as before (**Equations 23-24**); the changing sizes of these two populations resulting from the simulated intervention are assumed to change the overall mortality rate (**Equation 12**).

***Estimation of standard errors and confidence intervals***

To generate standard errors and confidence intervals around the lifetime-probability and life expectancy estimates, I considered as stochastic the parameter estimates generating the age-specific dementia incidence schedules (the fitted values of [α' β'] in **Equation 2**) and either prevalence (for Approach 1 – the fitted values of [α β] in **Equation 1**) or differential mortality (for Approach 2 – the fitted values for [α $\beta_1$ $\beta_2$ $\beta_3$] in **Equation 20**). Total mortality, derived from the SSA cohort life tables, was treated as deterministic (i.e. having zero variance) (Abatih, Van Oyen, Bossuyt, & Bruckers 2008; Loukine, Waters, Choi, & Ellison 2012), and the life table assumptions, such as linearity of survival within age intervals, were also considered not to contribute any additional variance.

For dementia incidence, I used the estimates of [α' β'] in **Equation 2**, along with their associated variance-covariance matrix, as the parameters of a bivariate Normal distribution to draw 1,000 independent values of [α' β'], generating 1,000 incidence schedules (Salomon, Mathers, Murray, & Ferguson 2001). Separately, I used an analogous procedure with the estimated parameters and variance-covariance matrix from **Equation 1** or **Equation 20** to generate 1,000 age schedules of prevalence (Approach 1)

or the mortality rate ratio between those with and those without dementia (Approach 2). Each incidence schedule was paired with one schedule of the second parameter (prevalence or mortality rate ratios) and run through the life table operations, producing 1,000 lifetime-probability and expectancy estimates. In figures, the median of the 1,000 estimates is shown as the point estimate, and the $2.5^{th}$ and $97.5^{th}$ percentiles are shown as the confidence bounds (Fishman 2015; Mooney 1997; Salomon et al. 2001). Tables show the means and standard errors (square roots of variances) of the 1,000 estimates. A deterministic result using the estimated parameters of incidence, prevalence, and differential mortality as fixed quantities will also be presented.

Parameters from **Equations 1, 2, and 20** were estimated using Stata version 14 (StataCorp, College Station, TX), using first-order Taylor Series linearization for variance estimation with the *svy* routine (Heeringa et al. 2009). Random sampling for the estimation of standard errors was conducted in R using the *mvrnorm* command in the MASS package (Venables & Ripley 2002), and life table operations were conducted using base R (R Core Team, Vienna, Austria). The HRS and ADAMS data are available to the public after a registration procedure (Health and Retirement Study 2013).

## 2.3 RESULTS

There were 308 cases of dementia at baseline out of 856 unweighted sample members. All baseline sample members were at risk of death, generating 519 deaths in 3,520 person-years at risk. Among the 456 individuals without dementia at baseline who were followed longitudinally, 106 developed dementia in 2,142 person-years at risk. The

estimates of the regression parameters in **Equations 1, 2, and 20**, along with the variance-covariance matrices used to sample the 1,000 simulated incidence, prevalence, and mortality rate-ratio schedules for the confidence intervals, are shown in **Appendix A.2.1**.

**Table 2-1** shows the estimated age schedule of dementia prevalence, which aligns closely with published estimates from ADAMS (Plassman et al. 2007), at 6.1% for age 75, 22.7% for age 85, and 38.4% at age 90. The fitted age-specific dementia incidence rates are also shown in Table 1, with 16 new cases per 1,000 person-years at age 75, 37 new cases per 1,000 person-years at age 85, and 56 to 86 new cases per 1,000 person-years at ages 90-95. The incidence rates shown here are close to those from the previous ADAMS incidence study, which reported incidence rates of 18.9 (95% CI: 10.1, 27.8) new cases per 1,000 person-years for ages below 80, 42.2 (95% CI: 26.0, 58.5) for ages 80 to 89, and 82.1 (95% CI: 39.9, 124.3) for ages 90 and above (Plassman et al. 2011). The small differences between the current incidence estimates and those from the prior study are attributable to the parametric model fit to the data in this study.

The age pattern of mortality rate ratios shown in Table 2-1, showing a rapid decline in differential mortality with age, is largely consistent with that found in other, non-national and non-U.S. samples; however, as mentioned previously, the pace of decline of differential mortality with age varies widely in the literature (Guehne et al. 2005; Ostbye et al. 1999; Tschanz et al. 2004; Villarejo et al. 2011). The estimated mortality rate ratio (RR) at age 70 of 13.3 is highly unreliable, as shown by its high standard error, arising from the low number of deaths at age 70; the true RR is probably

not quite that high. Since the prevalence of dementia is very low at age 70, results are not sensitive to the RR estimate at that age. The RR implied by estimated incidence, baseline prevalence, and the stationary population assumption is around 1 for ages above 85, which is plausible given the high level of mortality in the general population at those ages. The close correspondence between the point estimates using stochastic inputs (Table 2-1) and the deterministic estimates (**Appendix Table A.2.2a**), with the exception of the highly unreliable age-70 estimate of RR, suggest that the resampling strategy used to generate the stochastic estimates was on target.

Using Approach 1, **Table 2-2** shows calculations of the various life table quantities of interest: total life expectancy, dementia-free life expectancy and life expectancy with dementia (for a randomly chosen person in the population), and, of most interest, the probability that a currently dementia-free person will develop dementia later in life. It also shows conditional dementia-free life expectancy (DFLE') – dementia-free life expectancy for a dementia-free person at the given ages. About 23.7% (SE: 2.9%) of dementia-free 70 year old males are expected to develop dementia later in the course of life, compared to 31.8% (SE: 3.6%) of dementia-free females age 70. The lifetime probability remains roughly constant with age for males, meaning the force of dementia incidence increases about as quickly as the force of male mortality. For females, lifetime probability declines to 25.2% at age 95, indicating that the force of female mortality increases more quickly than the force of dementia incidence. Lifetime probability is higher for females than males because females have lower overall mortality, the only input that varies by sex. **Appendix Table 2.2.2b** again shows concordance between the

stochastic point estimates of expectancies and lifetime risk and the deterministic calculations.

For both sexes at ages 70 to 85, the vast majority of remaining life for a randomly chosen person is expected to be dementia-free, as shown by the high ratio of DFLE to DLE at these ages. However, at the oldest ages, 90 and above, DFLE and DLE are about equal – with DLE even greater than DFLE at age 95 – because mortality, incidence, and prevalence are all high at these ages. Dementia-free life expectancy for a dementia-free person (DFLE') is slightly higher than unconditional DFLE because the latter includes people with dementia, whose DFLE is zero. The gap between DFLE' and DFLE widens with age because the prevalence of dementia rises with age.

**Table 2-3** (with **Appendix Table A.2.2c** for deterministic quantities) shows the same quantities as Table 2-1 – prevalence, incidence, and the mortality rate ratio (RR) – but uses estimated incidence and RR to infer prevalence (i.e. Approach 2), rather than using estimated incidence and prevalence to infer RR. For all except the youngest and oldest ages, estimated RR is higher than that implied by stationary-population incidence and prevalence in Table 2-1. For example, 80-year-olds with dementia are estimated to die at 4.6 times the rate of 80-year-olds without dementia, whereas the stationary population implied a multiplier of just 1.8 times. A higher estimated RR implies a lower prevalence, also shown in Table 2-3, because exits from the population with dementia (relative to the population without) occur more quickly when RR is higher. **Figure 2-1** shows that the prevalence series implied by the estimated RRs is not even within the confidence bands of the prevalence from baseline ADAMS for ages 80 and above.

There are several possible explanations for the discordant results shown in **Figure 2-1**. If the estimated RRs are correct, then the figure implies a departure from stationarity, i.e. a change in dementia incidence rates, such that the rates that produced baseline prevalence were higher than the rates observed longitudinally in ADAMS. Another possible, though less likely, departure from stationarity consistent with **Figure 2-1** is that RR has increased over time, since a higher RR implies a lower prevalence. However, if we are confident that dementia incidence rates and differential mortality have not changed, then the figure implies a misspecification of the RR function (**Equations 21-22**). A final possibility is that the discrepancy arises from differences in the ascertainment of dementia status between the baseline study and the longitudinal follow-up study. Although the assessments of subjects were similar across waves of ADAMS, the baseline study by definition did not have access to the results of prior dementia examinations by the ADAMS team. To obtain a cognitive history of the subject (that is, measures of cognition prior to the baseline examination), the baseline study used medical records and interviews of knowledgeable informants. These methods have high reliability and validity (Langa et al. 2005; Plassman et al. 2007), but they are not the same as observing the person's cognitive performance in detail over time, as was done for the longitudinal ADAMS incidence sample.

When using the non-stationary approach, estimated lifetime risk of dementia is higher by about 3 to 5 percentage points than when using stationary population relations. For example, **Table 2-4** (with **Appendix Table A.2.2d** for deterministic quantities) shows that at age 70, the lifetime risk for males without dementia was 26.9% and for

females 34.7%. However, there is considerable overlap in the confidence intervals around the estimated-RR and stationary-population estimates of lifetime risk (**Figure 2-2**), suggesting statistical concordance in the results of the two approaches. In other words, the lifetime-risk estimates shown in **Table 2-2** are robust to the possible departures from stationarity implied by **Figure 2-1**. The higher estimate of differential mortality employed in Approach 2 means the competing risk of death without dementia is lower, raising lifetime risk of developing dementia.

Using the 1940 cohort life table rather than that of 1920 (with Approach 2) raises lifetime risk at all ages (**Table 2-5**). The increase is between 3 and 4 percentage points for both males and females. The probability that a dementia-free 70 year old male from this cohort develops dementia later in life is about 28.9%; for a dementia-free 70 year old female it is 34.9%. The increase in lifetime risk results from population-wide reductions in mortality between the two birth cohorts, reducing the competing risk of death and allowing a larger proportion of the population to survive to ages of high dementia incidence. The changes in the results based on the choice of an input life table do not negate the results for the older cohort, nor do they cast doubt on Approach 1, which requires only that incidence and differential mortality be constant over time. Rather, the 1940 results simply illustrate that individuals in younger, lower-mortality cohorts face higher age-specific lifetime risks of dementia than individuals in older, higher-mortality cohorts. The percentage increase in lifetime risk across the two cohorts is larger for males than females because females have lower mortality than males to begin with (a larger

base leads to smaller percentage change), and/or because mortality declined less for females than for males between these two cohorts (Preston & Wang 2006).

**Table 2-6** shows the lifetime-risk estimates under a series of alternative scenarios where dementia incidence is reduced, again using Approach 2. **Appendix A.2.3** shows the age pattern of incidence under the alternative scenarios. In the first two scenarios, an intervention delays the risk of dementia by one year; in Scenario 1 the intervention affects 50% of dementia-free 70-year-olds, and in Scenario 2 it affects 90%. The estimates for Scenarios 1 and 2 in **Table 2-6** indicate that this intervention would reduce lifetime risk at age 70 by only one to two percentage points, with similar reductions at older ages. The small difference between Scenarios 1 and 2 shows that the proportion of the age-70 population for which this intervention is effective has a small effect on lifetime risk estimates. Extending the reach from 50% to 90% of dementia-free 70-year-olds reduces lifetime risk by less than one percentage point.

A larger reduction in lifetime risk is achieved by an intervention that delays dementia onset by five years and reaches 50% of the dementia-free population age 70 (Scenario 3) – now the reduction is 3.7 percentage points for males and 4.5 for females. If this five-year delay affected 90% of dementia-free 70 year olds (Scenario 4), it would reduce lifetime risk at age 70 by 6.7 percentage points for males and 8.1 percentage points for females, a 25% reduction in lifetime risk for males and a 23% reduction for females. Similar reductions in lifetime risk are achieved by an intervention that reduces the rate of acceleration of dementia incidence with age, as in Scenario 5. This

intervention achieves a 5.6 percentage-point reduction in lifetime risk for males and 7.1 percentage-point reduction for females.

***Sensitivity analysis***

To test the sensitivity of the results to the choice of censoring subjects who survived ADAMS without dementia at their last assessment, a secondary analysis censored these subjects at the end of the ADAMS study period. This approach treats the two forms of not experiencing dementia onset – dying without developing dementia and surviving without developing dementia – in roughly the same way, following both groups until the end of ADAMS. This secondary approach increased the number of person-years at risk of death to 4,191 and the number of person-years at risk of dementia onset to 2,350. Because this sensitivity analysis only changed the treatment of subjects who were not observed as having experienced either event (dementia onset or death), estimates both of dementia incidence rates and of the mortality rate among the dementia-free population are lower than in the main analysis.

**Appendix Table A.2.5a** shows the estimated incidence rates under this alternative censoring strategy. They are in much less alignment than the main incidence rates (**Table 2-1**) with incidence rates previously published (Plassman et al. 2011). Furthermore, the relative risks implied by the incidence rates in Appendix Table 9 suggest the unlikely result that, for those aged 85 to 95, individuals with dementia died at a lower rate than individuals without dementia.

A comparison of **Appendix Table A.2.5b** to **Table 2-2** shows that lifetime risk of dementia is lower using the alternative censoring approach, by about 2 to 4 percentage points at younger ages and 5 to 6 percentage points at older ages. The lower incidence rates under this censoring strategy generate lower lifetime risks. Furthermore, the lower RR implied by imposing lower incidence rates on constant prevalence values raises competing risk of death without dementia.

**Appendix Table A.2.5c** also uses this alternative method of censoring but uses Approach 2, not assuming a stationary population and estimating relative risks of death directly from the data. Directly-estimated relative risks of death are higher using this method of censoring than when using the original method of censoring (**Table 2-3**) because this method of censoring counts more person-years lived without dementia, reducing the estimated mortality rate in the population without dementia and raising the RR. Incidence is again lower than in the main approach to censoring, with the same incidence estimates as in Appendix Table A.2.5a. In **Appendix Table A.2.5c,** the low estimates of incidence combine with the high estimates of RR to produce implausibly low estimates of dementia prevalence at old ages.

Lower incidence would imply lower lifetime risk of dementia than that estimated using the main approach to censoring, while higher RR would imply higher lifetime risk. **Appendix Table A.2.5d** shows that lifetime risk of dementia under this alternative censoring strategy is indeed lower than that estimated using the main censoring strategy (compare to **Table 2-4**), by about 2 to 4 percentage points for both sexes. **Appendix Table A.2.5e** is parallel to **Table 2-5**, showing results with Approach 2 (non-stationary

population) with the 1940 cohort life tables, but censoring those who survive ADAMS without dementia at the end of ADAMS rather than at their last assessment. As with the main results, using the 1940 life table raises lifetime risk of dementia by reducing the competing risk of mortality at ages of low dementia incidence, surviving a larger fraction of the population into ages of higher dementia incidence rates. The increase in lifetime risk due to the transition from 1920 to 1940 cohort mortality is about 3 to 4 percentage points at each age, similar in magnitude to the increase when using the main censoring strategy.

The discordance between the incidence rates in the sensitivity analysis and those in Plassman et al. 2011, and the implausibility of the implied RRs (**Appendix Table A.2.5a**) or the implied prevalence values (**Appendix Table A.2.5c**) cast doubt on the appropriateness of the censoring strategy employed in the sensitivity analysis and thus strengthen the validity of the approach taken in the main analysis.

## 2.4 DISCUSSION

This study provides, to my knowledge, the first nationally representative estimates of the lifetime probability of developing dementia in the U.S. These estimates suggest that about 23% to 27% of dementia-free 70 year-old males and about 31% to 35% of dementia-free 70-year-old females in the 1920 birth cohort will develop (or have developed) dementia before they die. For the 1940 birth cohort, these estimates rise to about 31% for males and 37% for females. The expected number of years that a randomly chosen individual at age 70 could expect to live with dementia is only about one to 1.5 years for males and two years for females, but given the high care needs of people with

dementia, this estimate still implies a large need for individuals and families to plan for a life stage with dementia.

A recent study, known as Adult Changes in Thought (Tom et al. 2015), reported dementia-free life expectancy for dementia-free cohort members (what I call DFLE') age 70, estimating 14.3 years for males and 15.7 years for females. My estimates were 11.1 years for males and 13.4 years for females (**Table 2-2**). The ACT population was clearly at lower baseline risk of mortality than the 1920 birth cohort in the U.S. as a whole, as the ACT cohort had a total life expectancy at age 70 of 16.0 years for males and 18.0 years for females, in contrast to the national cohort's 12.3 and 15.3 years. Since the ACT cohort had much longer life expectancy overall, it is not surprising that it also had longer conditional dementia-free life expectancy.

Another past study of an individual's lifetime risk of dementia that incorporated a competing-risks framework used Framingham data from 1975-1995 (Seshadri et al. 1997). It estimated that a dementia-free male age 65 had a 14.3% probability of developing dementia at some point in his remaining life, and a dementia-free female age 65 had 21.7% probability of developing dementia at some point in her remaining life (Seshadri & Wolf 2007).

There are several reasons why my estimates of the probability of developing dementia are considerably higher than the Framingham-based estimates. First, overall mortality during ADAMS was lower than overall mortality during Framingham. The age-standardized mortality rate (ASMR) in the U.S. population age-65+ in 2005, the middle of the ADAMS study period, was 4,804 deaths per 100,000 person-years lived; this rate

was much lower than the ASMR in Massachusetts for age 65+ in 1985, the middle of Framingham's study period, which was 5,679 deaths per 100,000 person-years lived (Centers for Disease Control and Prevention 2014). The comparison of the 1920 SSA cohort table to that of 1940 shows that lower mortality levels imply higher lifetime risks of dementia. Second, the mortality rate ratios I use could be higher than the differential mortality in the Framingham study, which was not specifically reported (Seshadri & Wolf 2007; Seshadri et al. 1997). A higher mortality rate ratio would reduce the competing risk of death without dementia and thus raise the estimated lifetime risk of developing dementia. Additionally, the authors of the Framingham study noted that because the Framingham sample was disproportionately white, it might not generalize to the U.S. population, due to known differences in incidence across racial and ethnic groups in the U.S. (Seshadri & Wolf 2007). ADAMS included a larger proportion of African American subjects, who are at higher risk of developing dementia at any given age (Plassman et al. 2011). Racial disparities in mortality also narrow with age (Eberstein et al. 2008; Preston, Elo, Rosenwaike, & Hill 1996). The higher incidence rates among African Americans could certainly contribute to higher estimates of lifetime risk of developing dementia. My estimates – especially those using the 1940 cohort life table from SSA – are closer to those estimated for a national sample from Canada, where the authors estimated that slightly over 40% of 70-year-olds in Canada would develop dementia before death (Carone et al. 2014).

A recent simulation study using ADAMS and HRS found that an intervention that delayed the onset of Alzheimer's Disease (AD) for five years would result in a 41%

lower prevalence of AD in 2050 among those aged 70+ than if onset had not been delayed (Zissimopoulos, Crimmins, & St.Clair 2014). Setting aside possible differences between AD and other dementias, a 41% decline in AD prevalence could be consistent with my estimate of a 25% decline in lifetime risk of dementia (Scenario 4 in **Table 2-6**), if the average duration of AD fell substantially as a result of the delayed onset of AD. In the literature, average duration of dementia among those who get it varies from 4.4 to 9.9 years depending on the age of diagnosis (Treves & Korczyn 2012). Additionally, since the average duration of AD is longer than that of vascular dementia (Treves & Korczyn 2012), an intervention reducing the duration of AD could have a larger effect on AD prevalence than an intervention reducing the duration of all types of dementia could have on all-dementia prevalence. Further research can use increment-decrement methods to estimate age-specific average duration of dementia under different possible interventions.

There are several other possible explanations for the difference between the potential reduction in prevalence obtained in the aforementioned simulation study and the potential reduction in lifetime risk estimated here. First, the prior study used a model incorporating changes in risk factors for dementia and death across real (not synthetic) cohorts, whereas I just used age and sex and a national life table. Second, they assumed that any intervention would be effective for 100% of the population at risk, while I only allowed up to 90% effectiveness. Additionally, it is unclear how they operationalized a five-year delay in onset of AD. They might have done so as I did, pushing age-specific incidence rates back five years and "filling in" the younger ages with lower rates consistent with the functional form of the model age pattern of dementia incidence. But

they might have set incidence to zero in the first five years and then continued five years of age later with the initial set of rates. Or perhaps they reestimated a different equation entirely. If they used a different operationalization of a "five year delay in onset," they could produce different results.

Another direction for future research is the estimation of quantities associated with family members of persons with dementia, such as the risk of having a parent who develops dementia. The estimates of lifetime risk presented here imply that informal care givers will face an increasing burden in the near future (Kasper et al. 2015). Because of the generally advanced age of persons with dementia, spouses are often not available to provide care, with daughters picking up the lion's share of the load (Friedman, Shih, Langa, & Hurd 2015). The lower parity of cohorts born in the 1940s and 1950s, relative to cohorts born in the 1920s and 1930s, implies that fewer aging Baby Boomers will have daughters who can take care of them than members of older generations have had (Human Fertility Database 2015). Research has found that middle-aged and young adults consistently underestimate their future need for personal care (Henning-Smith & Shippee 2015; Kemper, Komisar, & Alecxih 2005; Spillman & Lubitz 2002). The results shown here suggest that a large fraction of current and near-future elderly will develop dementia in their lifetimes, even if treatments delaying or reducing dementia risk become widespread.

**Table 2-1: Dementia Prevalence, Incidence, and Differential Mortality – Stationary Population Approach**

| | Fitted prevalence & incidence | | | | Implied | |
|---|---|---|---|---|---|---|
| Age | Prevalence | (SE) | Incidence | (SE) | RR | (SE) |
| 70 | 0.030 | (0.007) | 0.010 | (0.003) | 13.313 | (12.999) |
| 75 | 0.061 | (0.009) | 0.016 | (0.003) | 4.395 | (2.107) |
| 80 | 0.121 | (0.012) | 0.024 | (0.003) | 1.846 | (0.121) |
| 85 | 0.227 | (0.022) | 0.037 | (0.005) | 1.108 | (0.300) |
| 90 | 0.384 | (0.043) | 0.056 | (0.011) | 1.024 | (0.295) |
| 95 | 0.569 | (0.061) | 0.086 | (0.023) | 1.037 | (0.271) |
| 100 | 0.734 | (0.064) | 0.130 | (0.042) | 1.259 | (0.346) |

RR = Relative risk of death, with dementia vs. without dementia; SE = standard error.

Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For prevalence, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1920 birth cohort. Mortality by dementia status data from Health and Retirement Study (HRS), United States, 2001-2011.

Parametric models were fitted to prevalence and incidence data to generate single-year age-specific estimates. RR was implied by the fitted prevalence and incidence estimates along with stationary-population relations.

**Table 2-2: Life Cycle Quantities for Dementia – Stationary Approach**

A. Males

| Age | LE | DFLE | DLE | (SE) | Lifetime Risk | (SE) | DFLE' | (SE) |
|---|---|---|---|---|---|---|---|---|
| 70 | 12.31 | 10.83 | 1.48 | (0.129) | 0.237 | (0.029) | 11.14 | (0.119) |
| 75 | 9.65 | 8.04 | 1.61 | (0.142) | 0.233 | (0.027) | 8.53 | (0.134) |
| 80 | 7.26 | 5.54 | 1.72 | (0.162) | 0.226 | (0.032) | 6.25 | (0.143) |
| 85 | 5.20 | 3.43 | 1.76 | (0.182) | 0.215 | (0.041) | 4.37 | (0.135) |
| 90 | 3.64 | 1.91 | 1.74 | (0.187) | 0.208 | (0.052) | 3.01 | (0.109) |
| 95 | 2.61 | 0.96 | 1.64 | (0.163) | 0.213 | (0.067) | 2.15 | (0.078) |
| 100 | 2.02 | 0.47 | 1.55 | (0.123) | 0.246 | (0.089) | 1.71 | (0.039) |

B. Females

| Age | LE | DFLE | DLE | (SE) | Lifetime Risk | (SE) | DFLE' | (SE) |
|---|---|---|---|---|---|---|---|---|
| 70 | 15.25 | 13.17 | 2.08 | (0.190) | 0.318 | (0.036) | 13.41 | (0.191) |
| 75 | 11.91 | 9.77 | 2.14 | (0.205) | 0.304 | (0.037) | 10.19 | (0.205) |
| 80 | 8.91 | 6.74 | 2.16 | (0.227) | 0.286 | (0.044) | 7.41 | (0.209) |
| 85 | 6.37 | 4.25 | 2.12 | (0.249) | 0.267 | (0.054) | 5.17 | (0.190) |
| 90 | 4.42 | 2.41 | 2.00 | (0.251) | 0.253 | (0.067) | 3.53 | (0.149) |
| 95 | 3.10 | 1.26 | 1.83 | (0.220) | 0.252 | (0.081) | 2.47 | (0.102) |
| 100 | 2.32 | 0.63 | 1.69 | (0.167) | 0.277 | (0.102) | 1.90 | (0.061) |

LE = total life expectancy for a randomly chosen person in the population of given age, DFLE = Dementia-free life expectancy for a randomly chosen person in the population of given age, DLE = life expectancy with dementia for a randomly chosen person in the population of given age, Lifetime Risk = probability that a dementia-free person will develop dementia later in life, DFLE' = Dementia-free life expectancy for a dementia-free person of given age.

By construction, DFLE and DLE have the same standard error, and LE has zero variance.

Quantities calculated using fitted values of dementia incidence and prevalence and implied relative risk of death (with dementia vs. without) shown in Table 1.

Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For prevalence, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1920 birth cohort. Mortality by dementia status data from Health and Retirement Study (HRS), United States, 2001-2011.

**Table 2-3: Dementia prevalence, incidence, and differential mortality – non-stationary approach**

| Age | Implied Prevalence | (SE) | Estimated Incidence | (SE) | Estimated RR | (SE) |
|---|---|---|---|---|---|---|
| 70 | 0.030 | (0.173) | 0.010 | (0.003) | 8.86 | (3.344) |
| 75 | 0.058 | (0.241) | 0.016 | (0.003) | 6.37 | (1.761) |
| 80 | 0.093 | (0.304) | 0.024 | (0.003) | 4.63 | (0.896) |
| 85 | 0.137 | (0.371) | 0.037 | (0.005) | 3.41 | (0.513) |
| 90 | 0.196 | (0.443) | 0.056 | (0.011) | 2.54 | (0.447) |
| 95 | 0.286 | (0.535) | 0.086 | (0.023) | 1.91 | (0.479) |
| 100 | 0.453 | (0.673) | 0.130 | (0.042) | 1.48 | (0.472) |

RR = Relative risk of death, with dementia vs. without dementia

Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For relative risk of death, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1920 birth cohort. Mortality by dementia status data from Health and Retirement Study (HRS), United States, 2001-2011.

Parametric models were fitted to incidence and mortality data from ADAMS to generate single-year age-specific estimates. Prevalence in the cohort arises from the life table relations as described in the Methods section, under Approach 2.

**Table 2-4: Life Cycle Quantities – Non-Stationary Approach**

A. Males

| Age | LE | DFLE | DLE | (SE) | Lifetime Risk | (SE) | DFLE' | (SE) |
|---|---|---|---|---|---|---|---|---|
| 70 | 12.31 | 11.55 | 0.76 | (0.220) | 0.269 | (0.032) | 11.91 | (0.190) |
| 75 | 9.65 | 8.90 | 0.75 | (0.228) | 0.271 | (0.032) | 9.35 | (0.154) |
| 80 | 7.26 | 6.51 | 0.76 | (0.223) | 0.267 | (0.036) | 7.04 | (0.123) |
| 85 | 5.20 | 4.47 | 0.73 | (0.211) | 0.257 | (0.044) | 5.04 | (0.094) |
| 90 | 3.64 | 2.93 | 0.72 | (0.200) | 0.247 | (0.056) | 3.49 | (0.080) |
| 95 | 2.61 | 1.85 | 0.76 | (0.200) | 0.244 | (0.072) | 2.43 | (0.082) |
| 100 | 2.02 | 1.10 | 0.92 | (0.207) | 0.255 | (0.091) | 1.77 | (0.084) |

B. Females

| Age | LE | DFLE | DLE | (SE) | Lifetime Risk | (SE) | DFLE' | (SE) |
|---|---|---|---|---|---|---|---|---|
| 70 | 15.25 | 13.51 | 1.74 | (0.292) | 0.347 | (0.037) | 13.93 | (0.278) |
| 75 | 11.91 | 10.21 | 1.69 | (0.303) | 0.341 | (0.038) | 10.88 | (0.213) |
| 80 | 8.91 | 7.34 | 1.56 | (0.300) | 0.329 | (0.043) | 8.15 | (0.172) |
| 85 | 6.37 | 4.99 | 1.38 | (0.285) | 0.312 | (0.052) | 5.84 | (0.133) |
| 90 | 4.42 | 3.21 | 1.21 | (0.262) | 0.293 | (0.064) | 4.02 | (0.101) |
| 95 | 3.10 | 1.97 | 1.12 | (0.243) | 0.281 | (0.079) | 2.76 | (0.088) |
| 100 | 2.32 | 1.13 | 1.19 | (0.228) | 0.284 | (0.097) | 1.97 | (0.086) |

LE = total life expectancy for a randomly chosen person in the population of given age. DFLE = Dementia-free life expectancy for a randomly chosen person in the population of given age, DLE = life expectancy with dementia for a randomly chosen person in the population of given age, Lifetime Risk = probability that a dementia-free person will develop dementia later in life, DFLE' = Dementia-free life expectancy for a dementia-free person of given age.

By construction, DFLE and DLE have the same standard error, and LE has zero variance.

Quantities calculated using fitted values of dementia incidence and relative risk of death (with dementia vs. without) shown in Table 3.

Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For relative risk of death, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1920 birth cohort. Mortality by dementia status data from Health and Retirement Study (HRS), United States, 2001-2011.

**Table 2-5: Life Cycle Quantities Using 1940 Cohort Life Table – Non-Stationary Approach**

A. Males

| Age | LE | DFLE | DLE | (SE) | Lifetime Risk | (SE) | DFLE' | (SE) |
|---|---|---|---|---|---|---|---|---|
| 70 | 13.64 | 12.54 | 1.10 | (0.262) | 0.308 | (0.034) | 12.93 | (0.237) |
| 75 | 10.65 | 9.57 | 1.08 | (0.272) | 0.306 | (0.035) | 10.13 | (0.192) |
| 80 | 7.96 | 6.93 | 1.03 | (0.268) | 0.298 | (0.040) | 7.59 | (0.156) |
| 85 | 5.70 | 4.74 | 0.96 | (0.253) | 0.286 | (0.049) | 5.44 | (0.121) |
| 90 | 4.05 | 3.13 | 0.93 | (0.239) | 0.276 | (0.063) | 3.82 | (0.101) |
| 95 | 2.95 | 1.97 | 0.97 | (0.233) | 0.273 | (0.080) | 2.68 | (0.098) |
| 100 | 2.30 | 1.15 | 1.14 | (0.227) | 0.285 | (0.099) | 1.96 | (0.095) |

B. Females

| Age | LE | DFLE | DLE | (SE) | Lifetime Risk | (SE) | DFLE' | (SE) |
|---|---|---|---|---|---|---|---|---|
| 70 | 15.99 | 14.01 | 1.97 | (0.332) | 0.374 | (0.038) | 14.44 | (0.315) |
| 75 | 12.62 | 10.67 | 1.95 | (0.343) | 0.370 | (0.041) | 11.38 | (0.252) |
| 80 | 9.57 | 7.73 | 1.84 | (0.340) | 0.359 | (0.047) | 8.62 | (0.209) |
| 85 | 6.93 | 5.26 | 1.66 | (0.324) | 0.342 | (0.057) | 6.23 | (0.166) |
| 90 | 4.89 | 3.40 | 1.49 | (0.301) | 0.325 | (0.070) | 4.35 | (0.129) |
| 95 | 3.49 | 2.09 | 1.40 | (0.277) | 0.313 | (0.087) | 3.02 | (0.109) |
| 100 | 2.63 | 1.18 | 1.46 | (0.248) | 0.316 | (0.105) | 2.16 | (0.101) |

LE = total life expectancy for a randomly chosen person in the population of given age. DFLE = Dementia-free life expectancy for a randomly chosen person in the population of given age, DLE = life expectancy with dementia for a randomly chosen person in the population of given age, Lifetime Risk = probability that a dementia-free person will develop dementia later in life, DFLE' = Dementia-free life expectancy for a dementia-free person of given age.

By construction, DFLE and DLE have the same standard error, and LE has zero variance.

Quantities calculated using fitted values of dementia incidence and relative risk of death (with dementia vs. without) shown in Table 3. Mortality rates for total population come from Social Security Administration, 1940 birth cohort life tables.

Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For relative risk of death, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1940 birth cohort. Mortality by dementia status data from Health and Retirement Study (HRS), United States, 2001-2011.

**Table 2-6: Lifetime risk of dementia under intervention scenarios – non-stationary approach**

A. Males

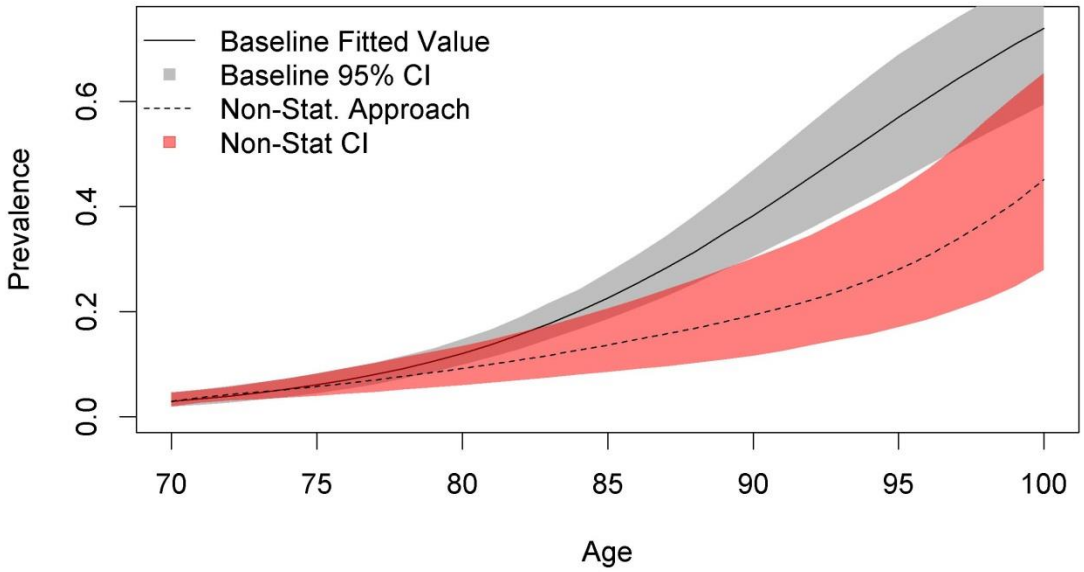
| | Scenario 1 | | Scenario 2 | | Scenario 3 | | Scenario 4 | | Scenario 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Age | Estimate | SE | Estimate | SE | Estimate | SE | Estimate | SE | Estimate | SE |
| 70 | 0.261 | (0.032) | 0.255 | (0.032) | 0.232 | (0.032) | 0.202 | (0.032) | 0.213 | (0.031) |
| 75 | 0.263 | (0.031) | 0.256 | (0.031) | 0.233 | (0.029) | 0.203 | (0.028) | 0.212 | (0.028) |
| 80 | 0.259 | (0.034) | 0.253 | (0.033) | 0.229 | (0.030) | 0.199 | (0.026) | 0.205 | (0.026) |
| 85 | 0.249 | (0.042) | 0.243 | (0.041) | 0.219 | (0.035) | 0.192 | (0.029) | 0.194 | (0.029) |
| 90 | 0.240 | (0.054) | 0.234 | (0.052) | 0.209 | (0.044) | 0.184 | (0.038) | 0.181 | (0.035) |
| 95 | 0.236 | (0.069) | 0.231 | (0.068) | 0.205 | (0.057) | 0.182 | (0.051) | 0.171 | (0.043) |
| 100 | 0.247 | (0.088) | 0.242 | (0.087) | 0.213 | (0.074) | 0.193 | (0.068) | 0.170 | (0.052) |

B. Females

| | Scenario 1 | | Scenario 2 | | Scenario 3 | | Scenario 4 | | Scenario 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Age | Estimate | SE | Estimate | SE | Estimate | SE | Estimate | SE | Estimate | SE |
| 70 | 0.338 | (0.036) | 0.330 | (0.036) | 0.302 | (0.035) | 0.266 | (0.034) | 0.276 | (0.034) |
| 75 | 0.331 | (0.037) | 0.324 | (0.036) | 0.296 | (0.034) | 0.260 | (0.032) | 0.268 | (0.031) |
| 80 | 0.320 | (0.042) | 0.312 | (0.041) | 0.284 | (0.036) | 0.250 | (0.032) | 0.255 | (0.031) |
| 85 | 0.303 | (0.050) | 0.296 | (0.049) | 0.268 | (0.042) | 0.236 | (0.036) | 0.237 | (0.035) |
| 90 | 0.285 | (0.062) | 0.278 | (0.060) | 0.251 | (0.052) | 0.222 | (0.045) | 0.217 | (0.041) |
| 95 | 0.273 | (0.077) | 0.267 | (0.075) | 0.239 | (0.064) | 0.214 | (0.058) | 0.200 | (0.049) |
| 100 | 0.277 | (0.094) | 0.271 | (0.093) | 0.240 | (0.080) | 0.218 | (0.074) | 0.192 | (0.057) |

Lifetime Risk = probability that a dementia-free person of given age will develop dementia later in life. Scenario 1: Dementia incidence delayed by 1 year, effective for 50% of dementia-free population age 70. Scenario 2: Dementia incidence delayed by 1 year, effective for 90% of dementia-free population age 70. Scenario 3: Dementia incidence delayed by 5 years, effective for 50% of dementia-free population age 70.
Scenario 4: Dementia incidence delayed by 5 years, effective for 90% of dementia-free population age 70. Scenario 5: Acceleration of dementia incidence with age reduced by 10%. Original incidence shown in Table 3.

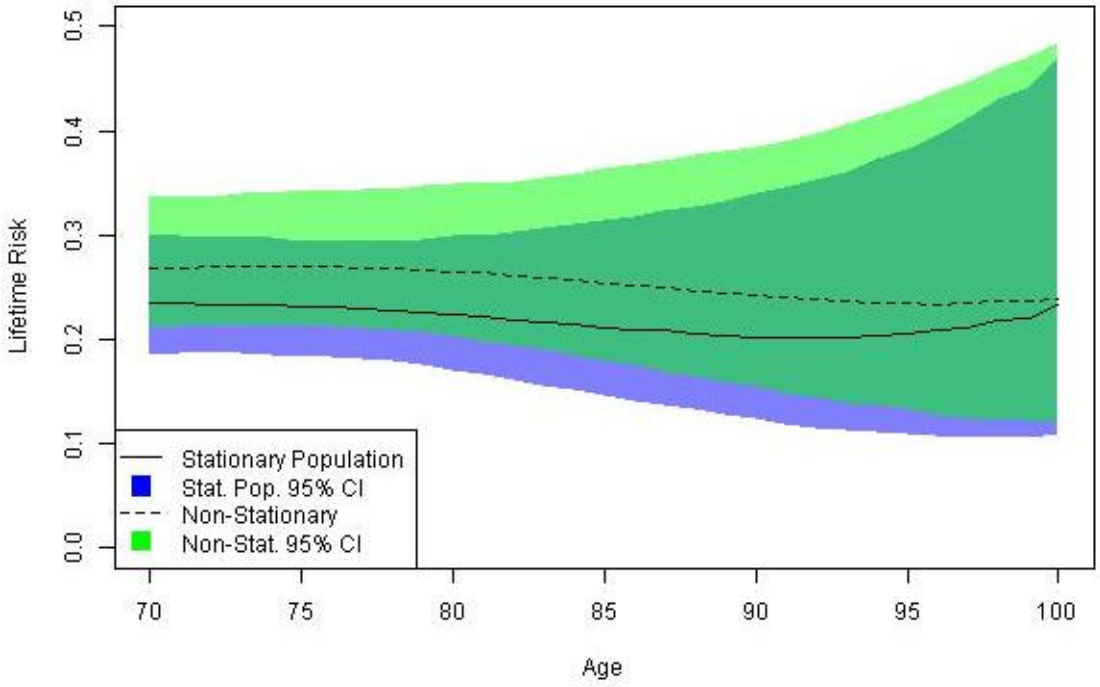
**Figure 2-1: Baseline fitted prevalence vs. prevalence implied by non-stationary approach**



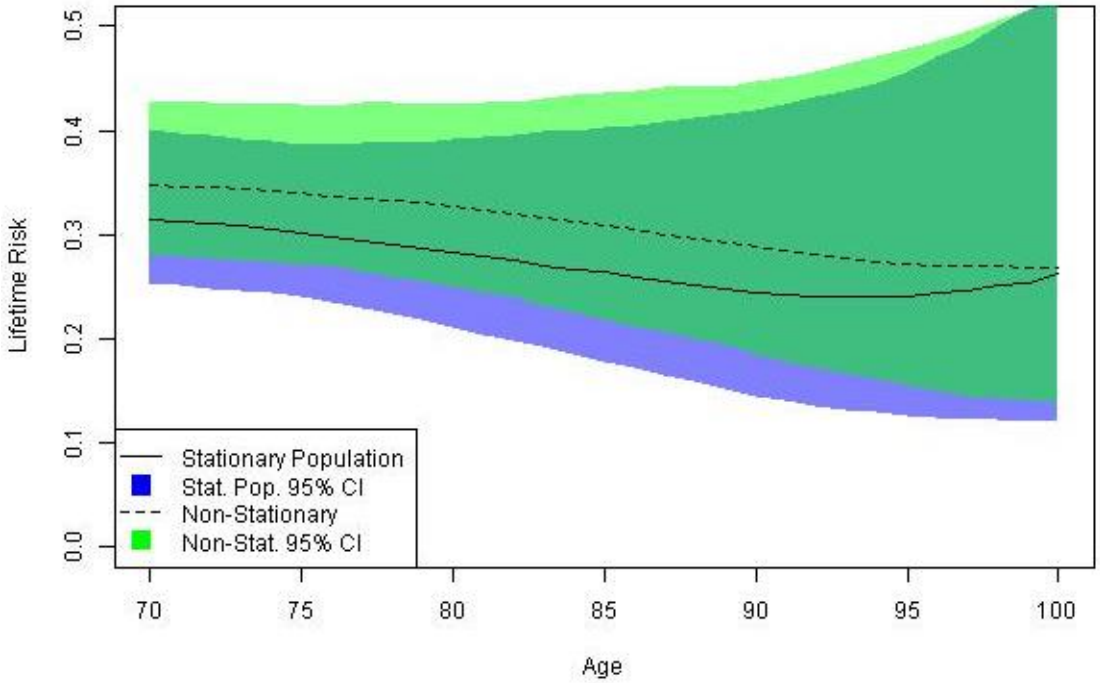

Notes: Baseline fitted prevalence from Aging, Demographics, and Memory Study (ADAMS) Wave A, United States, 2001-2003, n=856.
Estimated relative risks of death, used in Non-Stationary Approach, are from ADAMS and Health and Retirement Study, longitudinal follow-up 2001-2011, n=856 with 519 deaths. Prevalence implied by non-stationary approach arises from life table relations described in Methods section under Approach 2.

**Figure 2-2: Lifetime risk of dementia, stationary approach vs. non-stationary approach**

A) Males




B) Females




Notes: Lifetime risk = probability that a dementia-free person will develop dementia later in life. Stationary approach used baseline dementia prevalence and longitudinal dementia

incidence along with life-table relations and 1920 U.S. birth cohort life tables. Non-stationary approach used longitudinal dementia incidence and differential mortality estimates with the 1920 cohort life tables.

# 3 Incident diabetes and mobility limitations: reducing bias with risk-set matching

## 3.1 INTRODUCTION

Diabetes was estimated to account for $176B in annual medical costs and an additional $69B in lost productivity in the U.S. in 2012 (American Diabetes Association 2013). Population aging, increased age-specific diabetes incidence, and improved survival among people with diabetes are projected to cause the prevalence of diabetes in the U.S. among adults to increase from about 12% in 2010 (Y. J. Cheng et al. 2013) to between 21% and 33% by 2050 (Boyle et al. 2010). If the strong cross-sectional association between diabetes and disability (Wong et al. 2013) persists in the future, that projected prevalence of diabetes implies a massive burden of disability and medical costs on individuals, businesses, and government.

The cross-sectional associations between diabetes and physical functioning outcomes – the subject of most prior studies (Wong et al. 2013) – reflect causal pathways that can run in both directions: diabetes can cause mobility limitations and disability (Bianchi, Zuliani, & Volpato 2013), but mobility limitations among non-diabetic individuals can cause reductions in physical activity and other lifestyle changes that raise the risk of developing diabetes (Bardenheier et al. 2014). The latter pathway is a type of reverse causality that biases prevalence-based estimates of the association between diabetes and mobility limitations. Using incident, rather than prevalent, diabetes as the exposure of interest would reduce the strength of reverse causal pathways that plague previous studies of this topic. This paper will using risk-set matching, a type of

propensity score matching, to estimate the relationship between incident diabetes and subsequent mobility limitations, using data from the Health and Retirement Study (HRS).

The inability of prevalence-based estimates to identify the time-ordering of covariates can also create bias due to confounding. In prevalence-based studies, including longitudinal studies of the association between baseline diabetes prevalence and subsequent functional limitations (Chiu & Wray 2011; Cigolle et al. 2011), the values of common control variables such as body mass index (BMI), income, and comorbidities could actually have resulted from, rather than contributed to, the person's diabetes status (Bertoni & Goff 2011; Brown, Pagán, & Bastida 2005; Caruso, Silliman, Demissie, Greenfield, & Wagner 2000; C. Cheng et al. 2014). In estimating the association between prevalent diabetes and functional limitations, the investigator can either leave these covariates out of the model, raising the likelihood of confounding, or include them, running the risk of conditioning on a mediator. This study will provide a more accurate estimate of the relation between diabetes and functional limitation by controlling for factors known to precede diabetes diagnosis, including time-varying characteristics that change between baseline and diagnosis, and not for behaviors or conditions that follow it. Propensity-score methods are often preferred for reducing bias due to confounding (Rosenbaum & Rubin 1983; Rosenbaum 2010), and risk-set matching in particular is a preferred method for controlling for time-varying covariates when the exposure is time-varying (Rosenbaum 2010). A propensity-score based assessment has not been applied in the study of the relationship between incident diabetes and mobility limitations.

## 3.2 METHODS

### *Study population*

HRS is an ongoing longitudinal study representative of the U.S. population aged 50 and older, with biennial interviews collecting data on health status, demographics, health behavior, health care use, income and wealth (Health and Retirement Study 2013). Details on HRS are reported elsewhere (Heeringa & Connor 1995). A standardized data file made available by the RAND corporation was used (RAND 2011), confined to the original HRS cohort, born 1931 to 1941, and the "War Babies" cohort, born 1942 to 1947. The *baseline* observation for the original HRS cohort is 1992, and for the War Babies 1998. Each biennial interview of cohort members is called a *wave* and corresponds to a discrete date.

Response rates in the initial wave of each cohort were calculated as the number who responded divided by the number of eligible sample members. For subsequent waves, response rates were defined as the number who responded divided by the number of initial-wave respondents still alive. The original HRS cohort had an initial sample size (in 1992) of 12,652, representing an 81.6% overall response rate. Each subsequent wave had over an 85% response rate. The War Babies cohort had an initial sample size (in 1998) of 2,529, for a 69.9% response rate. Response rates in subsequent waves were always at least 87% (Health and Retirement Study 2011). Spouses of target respondents were included in the sample by HRS; I excluded spouses whose actual birth cohort did not align with their survey cohort, e.g. a spouse born in 1948 who was included in the War Babies survey cohort. Individual respondent weights are provided for each year a

subject is observed to make each sample cohort representative of the U.S. population born during the given years.

There were 12,230 subjects in the appropriate survey and birth cohorts. 501 were excluded because they did not respond to the baseline survey (1992 for HRS, 1998 for War Babies) but had been added to the dataset later. An additional 1,094 subjects reported having diabetes at baseline. The 987 subjects who provided fewer than three survey responses over the course of the study period were excluded, and two additional subjects had zero values for sampling weights, indicating their responses were valid for individual-level analysis (Health and Retirement Study 2010). The analytic sample size was thus 9,646. Time between first and last observation averaged 14.7 years and ranged from 4 to 18 years.

### *Measures*

Disability is a gap between personal capacity and environmental demand (Verbrugge & Jette 1994). Rather than being synonymous with diseases or impairments, disability is best viewed as a late stage in a pathway that begins with pathologies or diseases, such as diabetes, goes through impairments, such as neuropathy, then follows with functional limitations, such as pain or numbness in the lower extremities, and finally restrictions on a person's ability to function in society, such as the inability to get from place to place. Impairments affect organs or systems in the body; functional limitations refer to basic bodily or mental actions irrespective of the social context; and participation restrictions are about the person's interaction with society (Nagi 1976). The last category,

participation restrictions, is what the disabilities literature refers to as "disability" (Verbrugge & Jette 1994). Therefore, impairments and functional limitations are properly measured objectively, while disabilities are best understood subjectively. However, many activities, such as walking a specified distance or climbing stairs, the outcomes examined in this paper, fall on the boundaries between functional limitations and disabilities.

The outcome under study is lower-extremity mobility limitation, because of the biological link between diabetes and lower extremity impairments (Laditka & Laditka 2006). HRS asks, "Because of a health or memory problem, do you have any difficulty…?". The outcome variable can take values 0 through 5 because five lower-extremity mobility activities were measured: walking one block, walking several blocks, walking across a room, climbing one flight of stairs, and climbing several flights of stairs.

Diabetes status is determined by self-reported diagnosis. HRS asks, "Has a doctor ever told you that you have diabetes or high blood sugar?". Incident diabetes is defined as not reporting a diabetes diagnosis at one wave and reporting a diagnosis in the subsequent wave. An individual is considered to have diabetes for every wave subsequent to his/her initial report of diabetes diagnosis. There were 2,000 incident cases of diabetes during the years of data collection. The number of mobility limitations at the time of diagnosis is assumed to be the number of limitations at the last wave prior to diagnosis, to avoid possibly adjusting for limitations that follow from diabetes. A sensitivity analysis defined the number of mobility limitations at the time of diagnosis as the number of mobility limitations at the subject's first report of a diagnosis.

All covariates were self-reported. The following time-invariant covariates were used in the analysis: parents' years of education, self-rated health in childhood, gender, race/ethnicity (white, black, Hispanic), whether foreign born, and own years of education. The time-varying covariates were: age, marital status, region of residence (Northeast, Midwest, South, West), health insurance, employment status, income, wealth, smoking (current, former, never), BMI ($kg/m^2$), self-rated current health; the presence of high blood pressure, heart diseases, and arthritis; a history of stroke; and number of mobility limitations prior to diabetes diagnosis. A recent study showed that mobility limitations were associated with subsequent diabetes incidence, emphasizing the need to control for mobility limitations that precede it (Bardenheier et al. 2014). The rich set of socioeconomic variables, including parents' education and total wealth, distinguish this study from most of the literature on this topic (Wong et al. 2013).

Only four covariates had a large number of missing values: BMI (70 to 120 cases missing per wave), mother's education (845 missing cases, time-invariant), father's education (1,246 missing cases, time-invariant), and childhood health (355 missing cases, time-invariant – all individuals who were last observed in 1996). Missingness of a parent's education variable is likely correlated with own education, race, and income, which are likely correlated with both diabetes onset and disability accumulation. Single random imputation was used to assign values of parents' education to missing cases, under the assumption, known as "missing at random," that whether parents' education is missing is independent of its true value, conditional on observed variables. This is a much weaker assumption than the "missing completely at random" assumption employed by

other standard methods of dealing with missing data (Allison 2002). The parent's education variable was regressed onto sociodemographic covariates (age, gender, race/ethnicity, nativity, education, region, income). The observations missing parent's education take a value equal to the fitted value based on the regression plus a randomly generated error (Allison 2002).

For the missing BMI values, linear interpolation was used to assign a value based on the values the individual supplies in other waves: If the subject's BMI is 28 $kg/m^2$ in 1998 and 32 in 2002 and missing in 2000, their assigned BMI in 2000 is 30. All subjects had a valid baseline BMI value. Linear interpolation was unable to fill in BMI values for 534 person-observations. The individuals missing values for childhood health were all last observed in 1996, before the question was asked. In other words, missingness of the childhood health variable is a function of attrition from the study. Individuals last observed in 1996 were assigned the average childhood health value among individuals who died or were dropped from the study sample prior to 2010 but had nonmissing childhood health values; this average value was 1.80 with 1=best health, 5=worst health. All other covariates have fewer than 20 missing values per wave, negligible in a sample size of thousands.

***Matching method***

When an exposure might occur at different times, risk-set matching can be used to ensure that exposed individuals – those who report incident diabetes – and control individuals look similar prior to exposure while avoiding matching on characteristics

occurring after exposure (Rosenbaum 2010; Sylvestre, Huszti, & Hanley 2006). By matching individuals with similar characteristics up to the time of diabetes diagnosis, risk-set matching ensures that exposed and control units have been observed for the same amount of time, respecting the temporal structure of the data and avoiding so-called "immortal time bias" (Sylvestre et al. 2006), in which assignment to exposed or control groups requires using information obtained after the exposure. This bias applies when treated units have a built-in health advantage because they must have survived up to the point of treatment, whereas control units could have died at any time, or, as in the case of this study, when *control* units have a built-in health advantage because they must have survived until the end of the study period without exposure (Rosenbaum 2010).

A time-dependent propensity score was constructed using a discrete-time hazard model fit with logistic regression (Singer & Willett 1993). The propensity score is the predicted probability that the individual will develop diabetes in the given period, conditional on the values of the person's covariates. The observation period for the estimation of the propensity scores extends from baseline until the individual develops diabetes or is censored, whichever comes first.

Individuals are matched to minimize the total within-pair distance between exposed and control individuals, with "distance" defined as the rank-based Mahalanobis distance with a propensity-score caliper penalty function (Rosenbaum 2010; Silber et al. 2009). The variables included in the calculation of the Mahalanobis distance were the time-invariant covariates, the time-varying covariates measured at the wave prior to initial report of diabetes, and baseline BMI and self-rated health. Matching based on the

Mahalanobis distance minimizes the *total* within-pair distance between the exposed group's covariate matrix and the control group's covariate matrix, while the propensity score-based penalty ensures that the distance between any two matched individuals within a *single pair* is not too large (Rosenbaum 2010). Covariate balance is assessed by comparing standardized exposed-control differences in covariates, with standardized differences below 0.1 considered good balance and below 0.2 considered acceptable (Rosenbaum & Rubin 1985; Silber et al. 2001), and with significance tests for differences in means.

To be matched as an exposed unit, an individual needed to provide complete data – a full vector of independent variables and diabetes status – for both the first wave where diabetes is reported (i.e. when diabetes incidence is ascertained) and the wave immediately preceding it (when covariates are measured). If some information was missing, and especially if the individual had missing data for the wave immediately preceding the first report of diabetes, and thus the period of onset of their diabetes was uncertain, the individual was only eligible to be matched as a control unit. (See (Lu 2005) for details about data requirements in risk-set matching.)

At each wave, each incident diabetes case is matched to a diabetes-free individual. Cases are matched to controls separately for the original HRS and War Babies cohorts. Each control individual can only be matched to one case; see **Figure 3-1** for an illustration of the matching procedure. The matching was implemented using the *pairmatch* command in the *optmatch* package (Hansen & Klopfer 2006) in R (R Core Team, Vienna, Austria).

### *Analysis of outcome*

The dependent variable is first defined as the difference between an individual's mobility limitations at censoring and the individual's average number of mobility limitations prior to diabetes diagnosis (Lu 2005). Censoring time is the last wave in which both individuals in the pair were observed, ensuring the same length of follow-up within pairs. A one-sided Wilcoxon signed rank test is applied to the dependent variable to test the null hypothesis of no effect of incident diabetes on mobility limitations, and an asymmetric confidence interval is constructed by inverting the Wilcoxon test (Rosenbaum 2003). A rejection of the null hypothesis would imply the following: Under the assumption of no unmeasured confounding, incident diabetes increases the number of mobility limitations a person accumulates.

Then, a right-censored Poisson model was fit (Winkelmann 2003), with the dependent variable defined as the number of limitations at censoring time, accounting for length of follow-up. The independent variables are an indicator exposure status and the matched pair's total number of mobility limitations at censoring time (Lachin 2011). Because the matched pairs in this regression model are made using risk-set matching, the regression retains the useful properties of risk-set matching: time ordering is respected, and covariates are balanced both at baseline and just prior to exposure. The exponentiated coefficient, or rate ratio, for the exposure variable estimates the expected proportionate number of mobility limitations at study exit for a person who develops diabetes, relative

to a matched individual who does not. The model was fit using the *rcpoisson* command (Raciborski 2011) in Stata 13.1 (StataCorp, College Station, TX).

## 3.3 RESULTS

### *Descriptive statistics*

The average age of participants who reported no diabetes at study entry was 54.48 (SD: 3.04) and average number of mobility difficulties was 0.78 (SD: 1.22) (**Table 3-1**).

A comparison of individuals excluded from the analytic sample, due to having diabetes at baseline, to those included (**Appendix Table A.3.1a**) reveals that the excluded individuals tended to be more socioeconomically disadvantaged and less healthy at baseline than included individuals. Excluded individuals were more likely to be high school dropouts and to have had a history of smoking. They were less wealthy on average, and they reported more chronic conditions (high blood pressure, arthritis, and stroke) and more disabilities. Individuals who had diabetes at baseline had an average BMI more than 3 $kg/m^2$ higher than included individuals, over twice the prevalence of high blood pressure, and over four times the prevalence of having had a stroke.

**Appendix Table A.3.1b** shows the progression of each cohort through time. I observe 1,699 incident cases of diabetes in the HRS cohort and 301 in the War Babies cohort. In each cohort, the number of incident cases (and the implicit incidence rate, though this rate is subject to assumptions about mortality and attrition between the two data collection times) generally rise with age, but not monotonically. The number of

incident cases observed in a given wave is the maximum number of individuals who can be matched as "exposed" units for the period (interval) that precedes the given wave.

***Covariate balance***

**Table 3-2** shows the results of the matching procedure. 1,602 individuals who developed diabetes were matched to 1,602 controls who had not been diagnosed with diabetes as of the wave of matching. There were 136 individuals who reported diabetes during the study period but whose period of onset was uncertain, of which 9 were matched as controls and 127 were left unmatched. Additionally, to be included in analysis of the outcome, both units in a matched pair had to provide nonmissing data on mobility limitations at baseline, at the last wave prior to first diabetes report, and at censoring time. Since censoring time is defined as the last wave in which *both individuals* in a matched pair are observed, censoring time is unknown prior to matching. That raises the possibility that some matched pairs will have to be dropped following matching, if they have missing values for mobility limitations at censoring time. Nine such pairs were matched and then dropped, leaving 1,593 case-control pairs for analysis.

An approach to estimating the relation between incident diabetes and mobility limitations that does not use risk-set matching might take a sample of individuals who are non-diabetic at baseline and compare those who developed diabetes at any point during data collection to those who never developed diabetes. In the non-risk set matching approach, the difference in the mean number of mobility limitations at last wave prior to incident diabetes between the exposed group and the control group was 0.446 of a

limitation, or 62.7% of the mean in the never-diabetic group (**Table 3-2**). Comparisons of subsequent mobility limitations using the non-risk set matching approach would not control for this pre-exposure difference. An estimate of the diabetes-mobility limitation relationship using this approach would thus be biased upwards. In contrast, when using risk-set matching, the exposed-control group difference in means was 0.112 of a limitation, or 10.9% of the control group mean (**Table 3-2**), and in standard deviation units, the difference declined from 0.239 to 0.060.

**Table 3-2** also shows that, without matching, the exposed and control groups were statistically significantly different in terms of other key covariates, including BMI at baseline (exposed mean: 29.86 $kg/m^2$ vs. control mean: 26.31) and self-rated health just prior to diagnosis (3.031 vs. 2.568 where 1=excellent health). In all cases, the covariate differences between exposed and control groups prior to matching imply that the exposed group was both more likely to develop diabetes (see **Appendix Table A.3.2a**) and more likely to experience mobility limitation even without diabetes. Thus, a failure to balance exposed and control groups on these variables would bias estimates of the diabetes-mobility limitation relationship upward. **Table 3-2** shows that these variables were balanced after risk-set matching; for example, the difference in mean BMI at baseline between exposed individuals and matched controls was only 0.25 $kg/m^2$. See **Appendix Table A.3.2b** for a complete accounting of covariate differences before and after matching.

***Outcome analysis***

**Table 3-3** tests the null hypothesis that there was no difference in the change in mobility limitations between exposed individuals and their matched controls. At the median, individuals who developed diabetes experienced a 0.25-unit larger increase in mobility limitations than their matched controls (sign-rank $P$=0.0001). A 0.25-unit larger increase among cases than controls is substantively important, considering that the average number of limitations reported at baseline was only 0.79. The average post-diagnosis increase in self-reported mobility limitations in the exposed group was 0.55 of a mobility limitation, and in the matched control group 0.37 of a limitation (sign-rank $P$=0.0001).

The estimated rate ratio of 1.249 (**Table 3-4**, column 2) implies that people in the sample who developed diabetes had, on average, 24.9% more mobility limitations at study exit than matched individuals who did not develop diabetes, when both individuals were followed for the same post-exposure period. A comparison of the two columns in Table 4 shows a similar association between diabetes and mobility limitations regardless of the total number of mobility limitations in the matched pair.

***Sensitivity Analyses and Subgroup Analyses***

To test the sensitivity of the results to the use of self-reported diagnosis, similar analysis was performed excluding matched pairs in which the control unit would later develop diabetes. Such controls are most likely to have had undiagnosed diabetes at the time of matching. Results were similar when dropping matched pairs in which the control individual would later be diagnosed with diabetes (271 pairs; **Appendix Table A.3.3a**).

**Appendix Table A.3.3b** shows that results were also robust to an alternative parameterization of the number of mobility limitations at diabetes diagnosis.

In **Appendix A.3.4**, the sensitivity of the results to possible unmeasured confounding are examined, following Rosenbaum (Rosenbaum 1987, 2002). Suppose there exists an unmeasured confounder that is positively associated with the development of mobility limitations, such as a genetic predisposition toward development of diabetes after age 50. Such a variable would have to raise the probability of developing diabetes by 15% for one subject in each matched pair in order to explain the associations discussed above.

Finally, examining the association between diabetes incidence and mobility limitations among obese individuals could shed light on the complex relationships among obesity, diabetes, and mobility limitations. On the one hand, obesity could exacerbate the disabling consequences of diabetes; on the other hand, if obesity on its own exerts a strong negative effect on mobility, then it could mute the estimated effect of diabetes on mobility limitations. To explore the role of obesity, I analyze the outcome for the 466 matched pairs in which both the exposed and control units were obese at baseline.

The average mobility decline in the exposed group was 0.63 of a limitation and in the control group was 0.43 of a limitation, close to the results found for all matched pairs (**Appendix Table A.3.5a**). Similarly, **Appendix Tables A.3.5b and A.3.5c** show that the nonparametric test of significance and sensitivity of the results to unmeasured confounding produce very similar results when looking only at pairs in which both units were obese at baseline as when looking at all the matched pairs: a significant difference

in mobility decline, an estimated median difference of 0.25 of a limitation, and results that are somewhat sensitive to unmeasured confounding.

## 3.4 DISCUSSION

This study uses risk-set matching to estimate the association between incident diabetes and subsequent mobility limitations. Individuals who developed diabetes experienced significantly larger increases in mobility limitations than matched controls; the median difference in the change in mobility limitations was 0.25 limitations, out of a maximum of five. This is a fairly large difference, given that the average number of mobility limitations at baseline was 0.78 **(Table 3-1)**. At study exit, people in the sample who developed diabetes reported an estimated 24.9% more limitations than were reported by matched controls. These results were robust to a number of alternative specifications.

The study reduces many of the potential biases found in prior estimates of the relationship between diabetes and physical function limitations. In particular, by including only individuals who were diabetes-free at baseline and examining incident diabetes, bias from unobserved underlying health is reduced and factors that precede diabetes diagnosis are controlled without "controlling" for factors that follow it. Results show that the incident-diabetes and matched-control groups were balanced on measured confounders prior to diabetes onset.

The increased risk of physical functioning difficulties associated with diabetes as estimated in this study is lower than previous estimates (Cigolle et al. 2011; Dhamoon, Moon, Paik, Sacco, & Elkind 2014; Gregg et al. 2002; Kalyani, Saudek, Brancati, &

Selvin 2010; P. G. Lee et al. 2013; Wong et al. 2013; Wray, Ofstedal, Langa, & Blaum 2005). One difference between this and past investigations is that the current study controlled for a richer set of background socioeconomic and health conditions than most of the literature has done (Haas 2008; Herd, Goesling, & House 2007; Wong & Gregg 2013). Another important difference is that this study focuses on incident diabetes as the measure of exposure while prior studies focus on prevalent diabetes. As discussed in the introduction, prevalence-based estimates incorporate the effects of diabetes on functioning difficulties and the effects of functioning difficulties on the risk of developing diabetes. This study avoids incorporating the latter pathway, which is a source of bias.

Prevalence-based estimates also incorporate the effect of the duration of diabetes on functioning limitations, while the incidence-based estimates presented in this paper do not account for duration (see **Appendix A.3.6**). Since the duration of diabetes likely increases the risk of limitations, due to the progression of peripheral neuropathy and accumulation of comorbidities (Bruce, Davis, & Davis 2005; Stenholm et al. 2014), the estimates presented here might thus be considered a "lower bound" of the overall effect of diabetes on physical functioning limitations.

The study is subject to several limitations. First, because diabetes status was ascertained based on self-report, undiagnosed individuals are treated as non-diabetic. An alternative specification (**Appendix Table A.3.3a**) that excluded pairs in which the control individual later developed diabetes suggests that this limitation might not be severe, but it does not address the possibility that exposed individuals might have

developed diabetes years before they were diagnosed. Second, although missing data were addressed using recommended techniques, they could still create selection bias in the matching process. Third, the focus on mobility limitations ignores the many other types of functional limitations to which diabetes could contribute, such as limited vision (Cigolle et al. 2011). Examining other outcomes would provide a more comprehensive picture of the association between diabetes and physical limitations and would likely increase the estimated magnitude of the association (Cigolle et al. 2011).

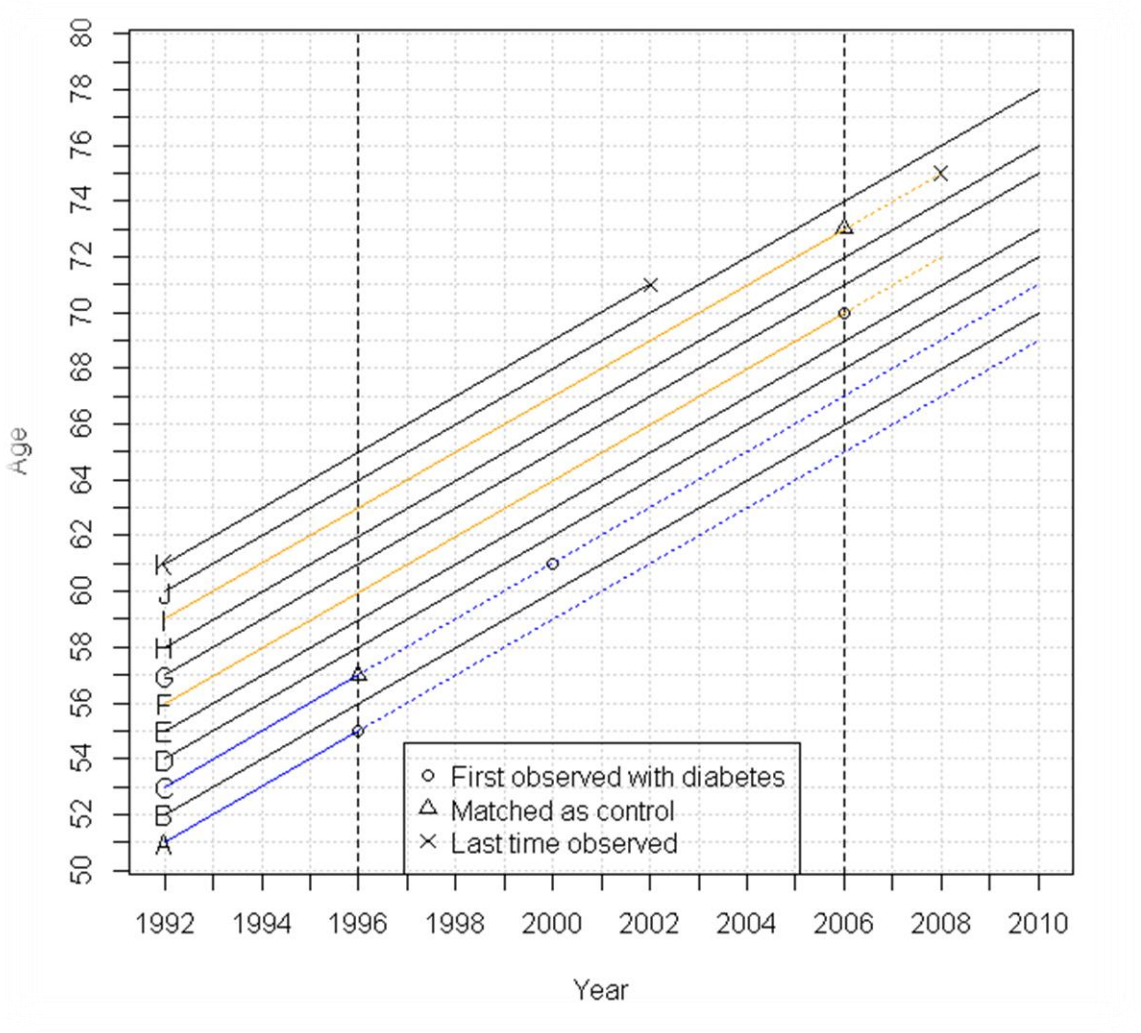
An assumption used throughout the analysis is that whether and when a person is censored from the study – through death or attrition – does not provide information on their accumulation of mobility limitations. Censoring by death may be informative; however, in theory, informative censoring could bias the results in either direction, and it is difficult to determine which direction is more likely. Finally, because the analysis focuses only on individuals born 1931 to 1947, it might not generalize to younger cohorts.

Although the results might be sensitive to unmeasured confounding (see **Appendix A.3.4**), the confounding would have to be of a fairly specific type. Unmeasured confounding that is positively associated with mobility limitation would have to raise the probability of developing diabetes, conditional not only on the host of early-life and current conditions this analysis controls for, but also on being diabetes-free at study entry. Those with diabetes at study entry were substantially sicker than the average sample individual on a range of measures. Unmeasured confounders that make people's underlying health poor – and thus increase the probability of both diabetes and

mobility limitation – are likely strongest among the people who already had diabetes at entry into the study. Because these individuals were excluded, general, unmeasured underlying health is less likely to confound the results.

As the U.S. population ages and the prevalence of chronic diseases such as diabetes increases, increasing our knowledge of the relationships between specific diseases and physical functioning difficulties will enhance our understanding of barriers to healthy aging among older Americans. The methods used in this paper reduce the bias in estimates of the effect of diabetes on increased mobility limitation. Given the high prevalence of mobility limitations and disabilities at older ages (Freedman et al. 2013), the rate ratios estimated in this paper amount to a large number of people whose mobility limitations result from diabetes. A randomized trial among people with diabetes showed that an intensive lifestyle intervention for weight loss improved physical functioning outcomes (Rejeski et al. 2014), implying considerable potential for improved diabetes management to reduce the burden of physical functioning limitations in the old-age population.

**Figure 3-1: Illustration of Risk-Set Matching**




Suppose there are eleven individuals in the cohort when the study begins in 1992, identified A through K. At the beginning of the study, they are aged 51 to 61 last birthday, respectively. In 1996, Subject A first reports having diabetes. At this point, Subject A is matched with someone in the *risk set*. The risk set consists of everyone who is still in the study, does not yet have diabetes, and has not yet been matched prior to 1996 (represented by solid life lines up to 1996). In this cohort, all subjects are in the risk set in 1996. Based on the similarity of their covariate matrices (characteristics observed

in 1992 and 1994), Subject C is matched as a control to Subject A. The pair is followed until the end of the study, represented by the dotted life lines. Subject C remains a control for Subject A even though Subject C reports having diabetes in 2000. The vertical dashed line at 1996 is the first "post-exposure" wave, i.e. the first wave after Subject A was diagnosed with diabetes. The change in disability from 1996 to 2010 for Subject A is compared to the change in disability from 1996 to 2010 for Subject C.

In 2006, subject F first reports having diabetes. Subject F is now matched with someone in the risk set, which consists of all the subjects with solid life lines up to 2006. Subjects A and C have already been matched, and subject K has died, so those three subjects are not in the risk set for 2006. Subject I is matched as a control to subject F. The pair is followed through 2008, which is the last time subject I is observed in the data set. The change in disability between from 2006 to 2008 for Subject F is compared to the change in disability from 2006 to 2008 for Subject I.

**Table 3-1: Weighted means (SDs) or percentages at study entry for individuals included in propensity score matching**

| Variable type | Variable | Mean (SD) or Percentage |
|---|---|---|
| Time-invariant | Mother's education (years) | 9.82 (3.59) |
| | Father's education (years) | 9.46 (3.94) |
| | Childhood health (1=excellent) | 1.76 (0.96) |
| | Female (%) | 50.55 |
| | Black (%) | 9.65 |
| | Hispanic (%) | 6.57 |
| | Foreign born (%) | 8.70 |
| | Education (years) | 12.77 (2.95) |
| Time-varying: demographic | Age | 54.48 (3.04) |
| | Married (%) [a] | 80.51 |
| | Not Married (%) [a] | 19.48 |
| | Northeast (%) | 19.42 |
| | Midwest (%) | 24.79 |
| | South (%) | 37.23 |
| | West (%) | 18.55 |
| Time-varying: economic | No health insurance (%) | 17.32 |
| | Works full-time (%) | 62.15 |
| | Works part-time (%) | 10.02 |
| | Unemployed (%) | 1.98 |
| | Retired (%) | 12.97 |
| | Not in labor force (%) | 9.35 |
| | On work disability (%) | 3.52 |
| | Income per household member ($) | 27,119 (40,401) |
| | Wealth ($) | 230,037 (489,000) |
| Time-varying: health | Smokes now (%) | 26.12 |
| | Ever smoked (%) | 62.23 |
| | BMI ($kg/m^2$) | 27.09 (5.07) |
| | Self-rated health (1=excellent) | 2.39 (1.13) |
| | High blood pressure (%) | 28.04 |
| | Heart problems (%) [b] | 7.70 |
| | Stroke (%) | 1.75 |
| | Arthritis (%) | 30.27 |
| | # Mobility limitations (max=5) | 0.78 (1.22) |
| | 0 (%) | 53.70 |
| | 1 (%) | 24.14 |
| | 2 (%) | 10.07 |

| | | |
|---|---|---|
| | 3 (%) | 5.18 |
| | 4 (%) | 4.15 |
| | 5 (%) | 2.72 |
| | Missing (%) | 0.04 |
| | # subjects | 9646 |

Notes: Health and Retirement Study, United States, 1992-2010.
[a] Marital status groups: married, partnered, widowed = "Married"; divorced, separated, never married = "Not married".
[b] Heart problems: heart attack, coronary heart disease, angina, congestive heart failure, or "other heart problems" (RAND 2011).

**Table 3-2: Covariate balance without vs. with risk-set matching, selected variables**

| When measured | Variable | No risk-set matching: Exposed mean | Control mean | Standardized diff [a] | *P* for diff [b] | With risk-set matching: Exposed mean | Control mean | Standardized diff [a] | *P* for diff [b] |
|---|---|---|---|---|---|---|---|---|---|
| Time-invariant | Female | 0.500 | 0.532 | -0.046 | 0.010 | 0.501 | 0.514 | -0.019 | 0.458 |
| | Own educ (yrs) | 11.7 | 12.5 | -0.181 | <0.001 | 11.8 | 11.9 | -0.010 | 0.914 |
| | Black | 0.215 | 0.140 | 0.138 | <0.001 | 0.209 | 0.180 | 0.053 | 0.040 |
| Study entry | Self-rated health [c] | 2.730 | 2.379 | 0.216 | <0.001 | 2.699 | 2.642 | 0.035 | 0.246 |
| | BMI ($kg/m^2$) [d] | 29.86 | 26.31 | 0.497 | <0.001 | 29.62 | 29.37 | 0.035 | 0.082 |
| | # mobility lims [e] | 1.203 | 0.822 | 0.202 | <0.001 | 1.200 | 1.082 | 0.063 | 0.017 |
| Observation prior to incident diabetes | Self-rated health [c] | 3.031 | 2.568 | 0.302 | <0.001 | 3.000 | 2.948 | 0.034 | 0.201 |
| | BMI ($kg/m^2$) [d] | 30.941 | 30.080 | 0.106 | <0.001 | 30.640 | 30.526 | 0.014 | 0.410 |
| | # mobility lims [e] | 1.156 | 0.710 | 0.239 | <0.001 | 1.138 | 1.026 | 0.060 | 0.057 |
| Number of individuals | | 2,000 | 7,646 | | | 1,602 | 1,602 | | |

Notes:

[a] Standardized difference = ((exposed group mean) – (control mean))/(Pooled standard deviation).

[b] P-values from Pearson chi-squared tests for binary variables (female, black), Wilcoxon rank-sum tests for numeric variables (e.g. BMI).

[c] Self rated health: 1=excellent, 5=poor.

[d] BMI = Body Mass Index.

[e] Maximum = 5.

**Table 3-3: Change in mobility limitations, exposed vs. matched control individuals**

| Parameter | Estimate |
|---|---|
| Hodges-Lehman estimate, 95% confidence interval [a] | 0.25 [0.0001,0.2501] |
| Sign-rank P value [b] | 0.0001 |
| Average change in # mobility limitations [c]: exposed group | 0.554 |
| Average change in # mobility limitations [c]: matched controls | 0.366 |
| P-value for difference [b] | 0.0001 |

Notes:
Num. pairs = 1,593.
[a] Hodges-Lehman estimate is the median of pairwise Walsh averages (approximately, the median within-pair difference). Confidence interval is asymmetric because of the highly discrete nature of the dependent variable.
[b] P-value from Wilcoxon signed-rank test for $H_0$: Change in mobility limitations among exposed = change in mobility limitations among matched controls.
[c] Change in mobility limitations = (# limitations at censoring) – ([# limitations just prior to exposed individual's diabetes diagnosis] + [# limitations at baseline])/2.

**Table 3-4: Rate Ratios from Censored Poisson Regression of Mobility Limitations onto Incident Diabetes**

Dependent variable = Number of mobility limitations at censoring (maximum=5, median=1) [a]

| | (1) | (2) |
|---|---|---|
| Incident diabetes | 1.220*** [1.122,1.328] | 1.249*** [1.149,1.357] |
| # limitations in pair | -- | 1.362*** [1.341,1.384] |
| *N* | 3186 | 3186 |

Notes:

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ (two-sided tests).

[a] Censoring time defined as last wave in which both observations in a matched pair were observed.

Model (1) regresses the dependent variable onto an indicator for being the "exposed" individual in the matched pair (i.e. the one who developed diabetes), accounting for length of follow-up. Model (2) adds a covariate for the total number of mobility limitations in the matched pair. Robust standard errors used.

## A.1. Appendices to Chapter 1 (Cohort Dynamics of Diabetes)

### A.1.1. Cohort Obesity Prevalence

The prevalence of obesity at age 25 in successive cohorts, used in our age/period/cohort model as a continuous variable, is shown in the following table:

**Table A.1.1 a: Prevalence of Obesity at age 25, by birth cohort**

| Birth Years | Percent |
|---|---|
| 1920-1924 | 2.00% |
| 1925-1929 | 2.58% |
| 1930-1934 | 3.34% |
| 1935-1939 | 3.46% |
| 1940-1944 | 3.96% |
| 1945-1949 | 4.85% |
| 1950-1954 | 5.99% |
| 1955-1959 | 6.99% |
| 1960-1964 | 8.47% |
| 1965-1969 | 11.32% |
| 1970-1974 | 14.52% |
| 1975-1979 | 15.92% |

Sources: Calculated from NHANES continuous waves 1999-2008 using the interview sample.

One limitation of our study is that we used retrospective data on height and weight to estimate trends in cohort obesity at age 25 for subsequent US birth cohorts. Recall data may be subject to errors of misreporting. However, prior research using longitudinal data found a relatively high degree of correspondence between recall and contemporaneously reported data on BMI (Perry, Byers, Mokdad, Serdula, & Williamson 1995). Validity of recall data over longer intervals of time has not been investigated. A second limitation is that we were not able to investigate cohort trends in obesity at younger ages because of lack of data on trends in childhood and adolescence.

### A.1.2. Prevalence Estimates and Confidence Intervals

Results displayed in Figures 1a and 1b are shown in more detail in the following tables.

**Table A.1.2 a: Age-Specific Prevalence Across Observation Periods**

| Period: | 1988-1994 | | 1999-2002 | | 2003-2006 | | 2007-2010 | |
|---|---|---|---|---|---|---|---|---|
| Age | Estimate | 95% CI | Estimate | 95% CI | Estimate | 95% CI | Estimate | 95% CI |
| 20-24 | 0.004397 | (0,0.009) | 0.004269 | (-0.001,0.009) | 0.00623 | (0.002,0.011) | 0.005288 | (-0.002,0.012) |
| 25-29 | 0.004864 | (0.001,0.009) | 0.017177 | (0.007,0.027) | 0.02294 | (0.01,0.035) | 0.021857 | (0.011,0.033) |
| 30-34 | 0.008792 | (0.004,0.014) | 0.026586 | (0.008,0.045) | 0.02416 | (0.013,0.035) | 0.028596 | (0.018,0.039) |
| 35-39 | 0.032287 | (0.015,0.05) | 0.029327 | (0.016,0.043) | 0.03726 | (0.022,0.052) | 0.039685 | (0.028,0.051) |
| 40-44 | 0.045436 | (0.03,0.06) | 0.046697 | (0.033,0.061) | 0.04710 | (0.03,0.064) | 0.052208 | (0.036,0.069) |
| 45-49 | 0.050596 | (0.036,0.066) | 0.070413 | (0.045,0.096) | 0.0677 | (0.048,0.087) | 0.085597 | (0.061,0.11) |
| 50-54 | 0.091078 | (0.063,0.119) | 0.096351 | (0.073,0.12) | 0.11732 | (0.095,0.14) | 0.1389 | (0.113,0.165) |
| 55-59 | 0.114349 | (0.091,0.138) | 0.114759 | (0.087,0.142) | 0.15295 | (0.112,0.194) | 0.156262 | (0.12,0.192) |
| 60-64 | 0.153817 | (0.129,0.178) | 0.178177 | (0.146,0.21) | 0.16539 | (0.138,0.193) | 0.192598 | (0.151,0.234) |
| 65-69 | 0.140782 | (0.109,0.172) | 0.193894 | (0.158,0.23) | 0.20029 | (0.164,0.237) | 0.264501 | (0.204,0.325) |
| 70-74 | 0.142179 | (0.112,0.173) | 0.160267 | (0.126,0.194) | 0.21140 | (0.175,0.247) | 0.248475 | (0.218,0.278) |
| 75-79 | 0.186321 | (0.148,0.224) | 0.160083 | (0.113,0.207) | 0.18344 | (0.145,0.222) | 0.230738 | (0.182,0.279) |

**Table A.1.2 b: Age-Specific Diabetes Prevalence by Birth Cohort**

| Cohort: | 1910-1919 | | 1920-1929 | | 1930-1939 | | 1940-1949 | |
|---|---|---|---|---|---|---|---|---|
| Age | Estimate | 95% CI | Estimate | 95% CI | Estimate | 95% CI | Estimate | 95% CI |
| 20-24 | | | | | | | | |
| 25-29 | | | | | | | | |
| 30-34 | | | | | | | | |
| 35-39 | | | | | | | | |
| 40-44 | | | | | | | 0.0613 | (0.031,0.091) |
| 45-49 | | | | | | | 0.0506 | (0.036,0.066) |
| 50-54 | | | | | 0.0743 | (0.048,0.1) | 0.1043 | (0.075,0.134) |
| 55-59 | | | | | 0.1143 | (0.091,0.138) | 0.1460 | (0.119,0.173) |
| 60-64 | | | 0.1596 | (0.128,0.191) | 0.1626 | (0.131,0.194) | 0.1814 | (0.158,0.205) |
| 65-69 | | | 0.1408 | (0.109,0.172) | 0.2059 | (0.178,0.234) | 0.2430 | (0.191,0.295) |
| 70-74 | 0.1337 | (0.098,0.17) | 0.1506 | (0.119,0.182) | 0.2190 | (0.198,0.24) | 0.2620 | (0.13,0.394) |
| 75-79 | 0.1863 | (0.148,0.224) | 0.1724 | (0.14,0.204) | 0.2246 | (0.179,0.27) | | |

(Table A.1.2b, continued)

| Cohort: | 1950-1959 | | 1960-1969 | | 1970-1979 | | 1980-1989 | |
|---|---|---|---|---|---|---|---|---|
| Age | Estimate | 95% CI | Estimate | 95% CI | Est. | 95% CI | Est. | 95% CI |
| 20-24 | | | 0.0079 | (0,0.016) | 0.0022 | (-0.001,0.006) | 0.0058 | (0.002,0.009) |
| 25-29 | | | 0.0049 | (0.001,0.009) | 0.0211 | (0.013,0.03) | 0.0201 | (0.009,0.031) |
| 30-34 | 0.0051 | (0.001,0.009) | 0.0203 | (0.01,0.031) | 0.0257 | (0.017,0.034) | 0.0364 | (-0.009,0.081) |
| 35-39 | 0.0323 | (0.015,0.05) | 0.0324 | (0.023,0.042) | 0.0403 | (0.029,0.052) | | |
| 40-44 | 0.0364 | (0.023,0.05) | 0.0507 | (0.04,0.061) | 0.0497 | (0.007,0.092) | | |
| 45-49 | 0.0733 | (0.056,0.09) | 0.0770 | (0.057,0.097) | | | | |
| 50-54 | 0.1229 | (0.106,0.14) | 0.1054 | (0.045,0.166) | | | | |
| 55-59 | 0.1406 | (0.109,0.172) | | | | | | |
| 60-64 | 0.1721 | (0.105,0.24) | | | | | | |
| 65-69 | | | | | | | | |
| 70-74 | | | | | | | | |
| 75-79 | | | | | | | | |

### A.1.3. Results of Models of Prevalence of Diabetes

Results displayed in Figures 2 and 3 are shown in more detail in the following tables.

**Table A.1.3 a: Results of Age-Cohort Model**

Dependent Variable = Log Diabetes Prevalence

| Indicator | Coefficient | SE | t-statistic | p-value | 95% CI | |
|---|---|---|---|---|---|---|
| Age 25-29 | 1.080695 | 0.292925 | 3.689325 | 0.001283 | 0.473206 | 1.688185 |
| Age 30-34 | 1.704125 | 0.314012 | 5.426944 | 1.89E-05 | 1.052904 | 2.355346 |
| Age 35-39 | 2.532282 | 0.328767 | 7.702351 | 1.1E-07 | 1.85046 | 3.214103 |
| Age 40-44 | 2.96014 | 0.342128 | 8.652146 | 1.57E-08 | 2.25061 | 3.669669 |
| Age 45-49 | 3.384705 | 0.367831 | 9.201791 | 5.36E-09 | 2.62187 | 4.14754 |
| Age 50-54 | 3.988511 | 0.395827 | 10.0764 | 1.05E-09 | 3.167616 | 4.809406 |
| Age 55-59 | 4.232604 | 0.430616 | 9.829178 | 1.65E-09 | 3.33956 | 5.125647 |
| Age 60-64 | 4.553761 | 0.435592 | 10.45419 | 5.35E-10 | 3.650399 | 5.457124 |
| Age 65-69 | 4.748697 | 0.46782 | 10.15069 | 9.19E-10 | 3.778497 | 5.718897 |
| Age 70-74 | 4.791991 | 0.494981 | 9.68117 | 2.17E-09 | 3.765464 | 5.818518 |
| Age 75-79 | 4.92001 | 0.536914 | 9.163504 | 5.77E-09 | 3.806519 | 6.0335 |
| 1920-29 cohort | 0.0491136 | 0.446574 | 0.109979 | 0.913423 | -0.87702 | 0.975251 |
| 1930-39 cohort | 0.3153463 | 0.443356 | 0.711272 | 0.48439 | -0.60412 | 1.234809 |
| 1940-49 cohort | 0.4939475 | 0.492615 | 1.002704 | 0.326905 | -0.52767 | 1.515569 |
| 1950-59 cohort | 0.5045705 | 0.527574 | 0.956399 | 0.349263 | -0.58955 | 1.598691 |
| 1960-69 cohort | 0.809644 | 0.554556 | 1.459987 | 0.158427 | -0.34043 | 1.959722 |
| 1970-79 cohort | 1.161477 | 0.583122 | 1.991825 | 0.058947 | -0.04784 | 2.370798 |
| 1980-89 cohort | 1.588968 | 0.630111 | 2.521726 | 0.019422 | 0.282197 | 2.895739 |
| constant | -6.687672 | 0.60934 | -10.9753 | 2.16E-10 | -7.95137 | -5.42398 |
| R-Squared | 0.9486 | | | | | |
| N | 41 | | | | | |

Age/Cohort Model: $ln(Y_{ia}) = \alpha + \beta_a X_a + \beta_i X_i$, where $Y_{ia}$ = the proportion of the population in cohort $i$ at age $a$ with diabetes, $X_a$ is a dummy variable indicating that the observation pertains to age $a$, and $X_i$ is a dummy variable indicating that the observation pertains to cohort $i$.

**Table A.1.3 b: Results of Age-Period Model**

Dependent Variable: Log Diabetes Prevalence

| Indicator | Coefficient | SE | t-statistic | p-value | 95% CI | |
|---|---|---|---|---|---|---|
| Age 25-29 | 0.8340861 | 0.177858 | 4.689627 | 4.59E-05 | 0.472232 | 1.19594 |
| Age 30-34 | 1.229262 | 0.176308 | 6.972242 | 5.69E-08 | 0.87056 | 1.587963 |
| Age 35-39 | 1.935714 | 0.176051 | 10.99518 | 1.43E-12 | 1.577535 | 2.293892 |
| Age 40-44 | 2.256934 | 0.173813 | 12.98482 | 1.61E-14 | 1.903309 | 2.61056 |
| Age 45-49 | 2.564451 | 0.181592 | 14.12202 | 1.52E-15 | 2.194998 | 2.933903 |
| Age 50-54 | 3.056156 | 0.183954 | 16.61367 | 1.36E-17 | 2.681898 | 3.430414 |
| Age 55-59 | 3.259276 | 0.195733 | 16.65169 | 1.27E-17 | 2.861055 | 3.657497 |
| Age 60-64 | 3.512932 | 0.177514 | 19.78963 | 7.07E-20 | 3.151777 | 3.874086 |
| Age 65-69 | 3.625526 | 0.187088 | 19.37873 | 1.34E-19 | 3.244892 | 4.006159 |
| Age 70-74 | 3.586429 | 0.189392 | 18.93653 | 2.7E-19 | 3.201108 | 3.97175 |
| Age 75-79 | 3.620085 | 0.210792 | 17.17377 | 5.08E-18 | 3.191226 | 4.048944 |
| 2001 NHANES | 0.2922673 | 0.108457 | 2.694766 | 0.010993 | 0.071609 | 0.512926 |
| 2005 NHANES | 0.4265295 | 0.108519 | 3.930467 | 0.00041 | 0.205746 | 0.647313 |
| 2009 NHANES | 0.5343021 | 0.100049 | 5.34043 | 6.76E-06 | 0.330752 | 0.737852 |
| constant | -5.571567 | 0.133038 | -41.8794 | 3.46E-30 | -5.84224 | -5.3009 |
| R-Squared | 0.9686 | | | | | |
| N | 48 | | | | | |

Age/Period model: $ln(Y_{ia}) = \alpha + \beta_a X_a + \beta_p X_p$, where $Y_{ia}$ = the proportion of the population in cohort *i* at age *a* with diabetes, $X_a$ is a dummy variable indicating that the observation pertains to age *a*, and $X_p$ is a dummy variable indicating that the observation pertains to period *p*.

**Table A.1.3 c: Results of Age-Period-Cohort Obesity Model**

Dependent Variable: Log Diabetes Prevalence

| Indicator | Coefficient | SE | t-statistic | p-value | 95% CI | |
|---|---|---|---|---|---|---|
| Age 25-29 | 0.9926029 | 0.155681 | 6.375895 | 5.71E-07 | 0.6742 | 1.311005 |
| Age 30-34 | 1.655445 | 0.18084 | 9.154213 | 4.71E-10 | 1.285587 | 2.025304 |
| Age 35-39 | 2.672352 | 0.22659 | 11.79378 | 1.38E-12 | 2.208924 | 3.13578 |
| Age 40-44 | 3.191595 | 0.25972 | 12.28859 | 5.08E-13 | 2.660407 | 3.722782 |
| Age 45-49 | 3.636119 | 0.288422 | 12.60695 | 2.71E-13 | 3.046231 | 4.226008 |
| Age 50-54 | 4.259973 | 0.314103 | 13.56235 | 4.38E-14 | 3.61756 | 4.902385 |
| Age 55-59 | 4.545092 | 0.333427 | 13.63146 | 3.86E-14 | 3.863158 | 5.227026 |
| Age 60-64 | 4.892977 | 0.345447 | 14.16418 | 1.46E-14 | 4.186459 | 5.599496 |
| Age 65-69 | 5.07219 | 0.360942 | 14.05265 | 1.79E-14 | 4.333981 | 5.810399 |
| Age 70-74 | 5.107939 | 0.392845 | 13.00242 | 1.26E-13 | 4.304481 | 5.911398 |
| Age 75-79 | 5.116603 | 0.416248 | 12.2922 | 5.04E-13 | 4.26528 | 5.967926 |
| 2001 | - | | | | | |
| NHANES | 0.0517601 | 0.119307 | -0.43384 | 0.667616 | -0.29577 | 0.192251 |
| 2005 | | | | | | |
| NHANES | 0.0782829 | 0.119932 | 0.652728 | 0.519073 | -0.16701 | 0.323571 |
| 2009 | | | | | | |
| NHANES | 0.0561693 | 0.142336 | 0.394624 | 0.696008 | -0.23494 | 0.347279 |
| obesity | 13.10013 | 2.828275 | 4.631844 | 7.05E-05 | 7.315658 | 18.8846 |
| constant | -7.107672 | 0.35859 | -19.8212 | 2.09E-18 | -7.84107 | -6.37427 |
| R-Squared | 0.9813 | | | | | |
| N | 45 | | | | | |

Age/Period/Cohort model: $ln(Y_{ia}) = \alpha + \beta_a X_a + \beta_p X_p + \gamma Coh\_ob$, where $Y_{ia}$, $X_a$, and $X_p$ are defined as in the Age/Period model and *Coh_ob* is a continuous variable representing the prevalence of obesity at age 25 in the cohort corresponding to the given age and period.

### A.1.4. Results when using Fasting Plasma Glucose as the Indicator of Diabetes

The figures below (A.1.4a- 4c) are analogous to Figures 1-3, 1-4, and 1-5 in the main manuscript. The difference is that in the figures below, diabetes is defined as fasting plasma glucose ≥ 126 mg/dL and not based on HbA1c. Data are from the National Health and Nutrition Examination Surveys (NHANES) II, III, and Continuous, United States, 1976-2010.

**Figure A.1.4 a: Prevalence as a multiple of 1900-1910 birth cohort prevalence**

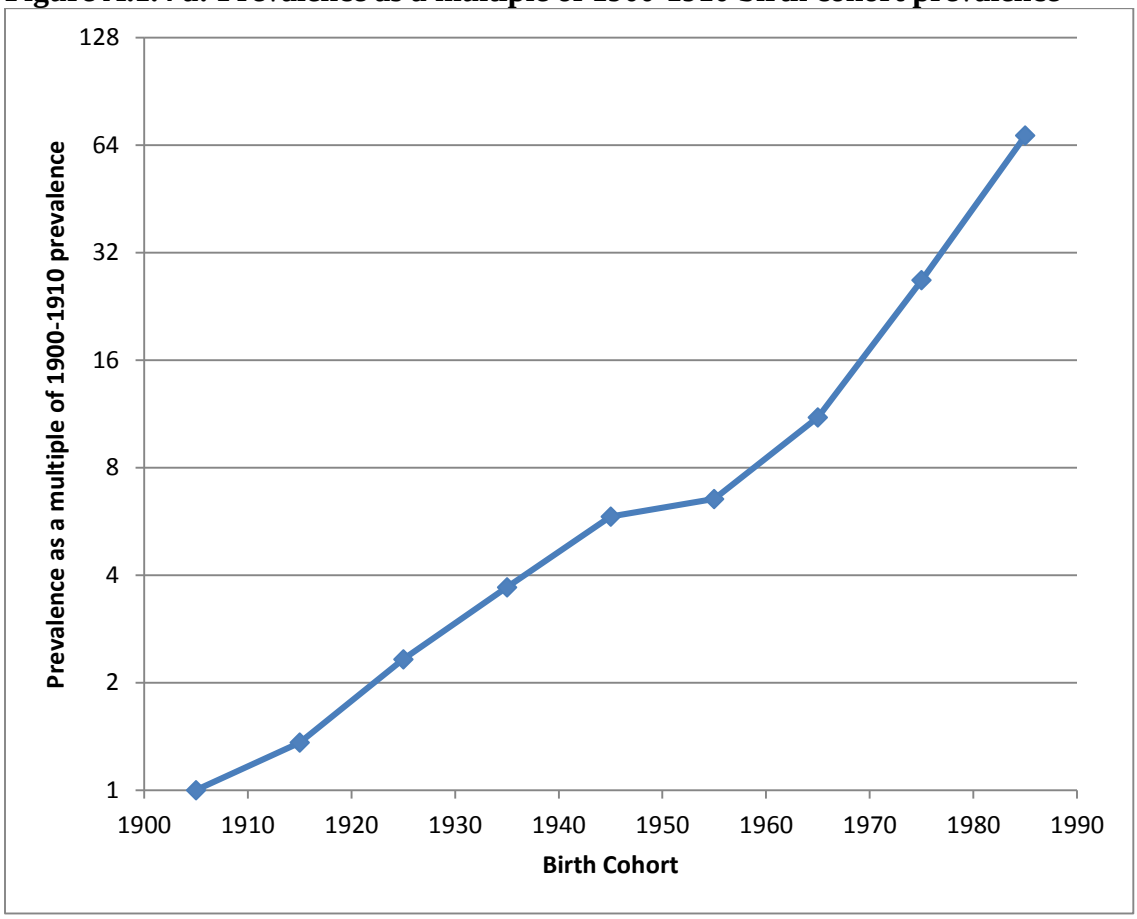


**Figure A.1.4 b: Age-adjusted diabetes prevalence across observation periods**

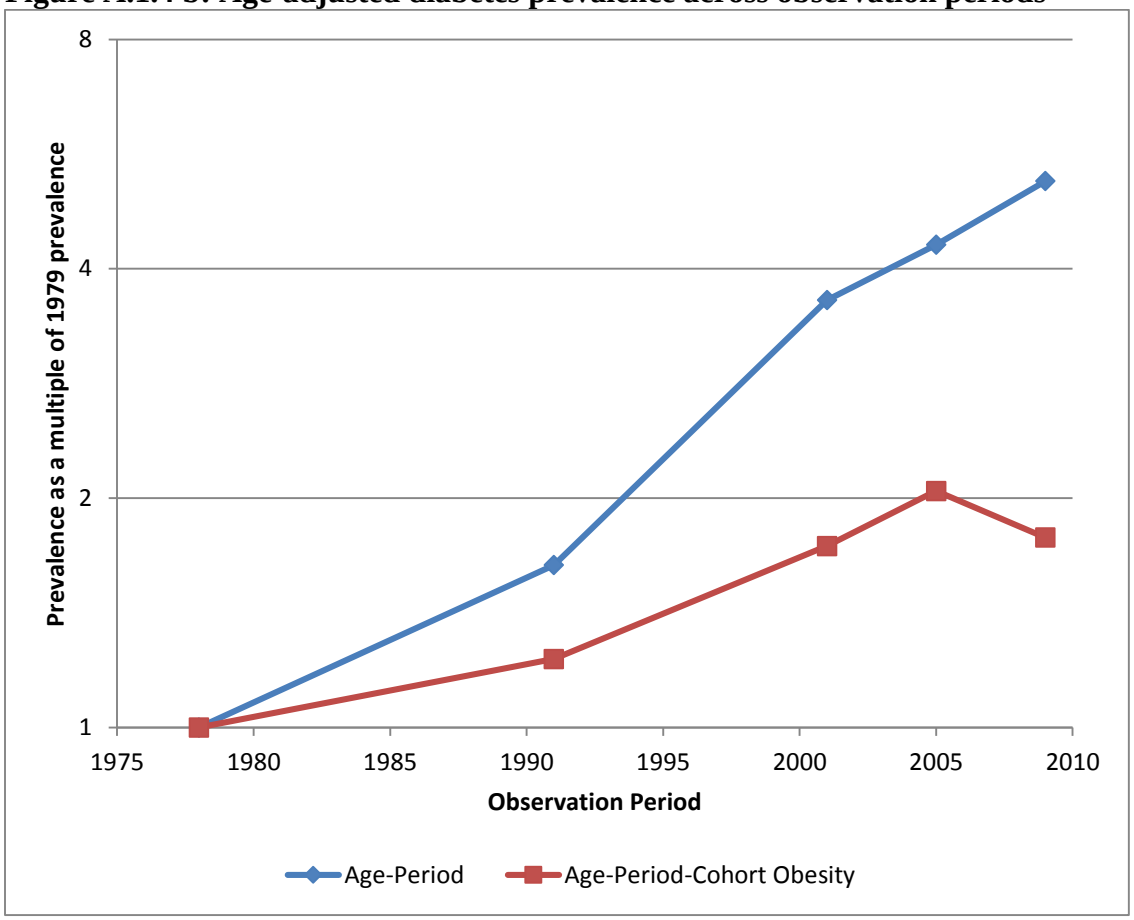
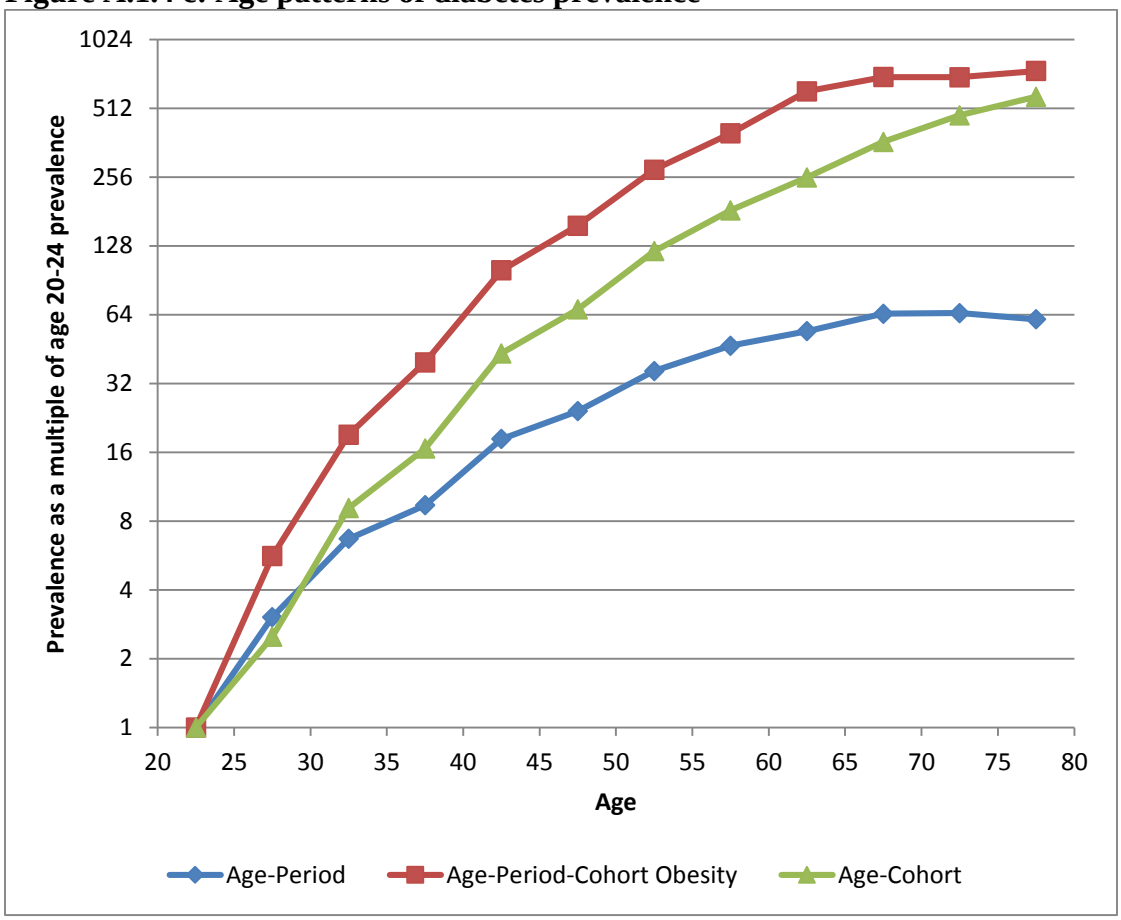


**Figure A.1.4 c: Age patterns of diabetes prevalence**



### A.1.5. Model Age Pattern of Diabetes Incidence

The model age pattern of diabetes incidence (new cases per person-year without diabetes) shown in Figure 4 in the main text is shown in detail (values for 1950-59 birth cohort) in the following table:

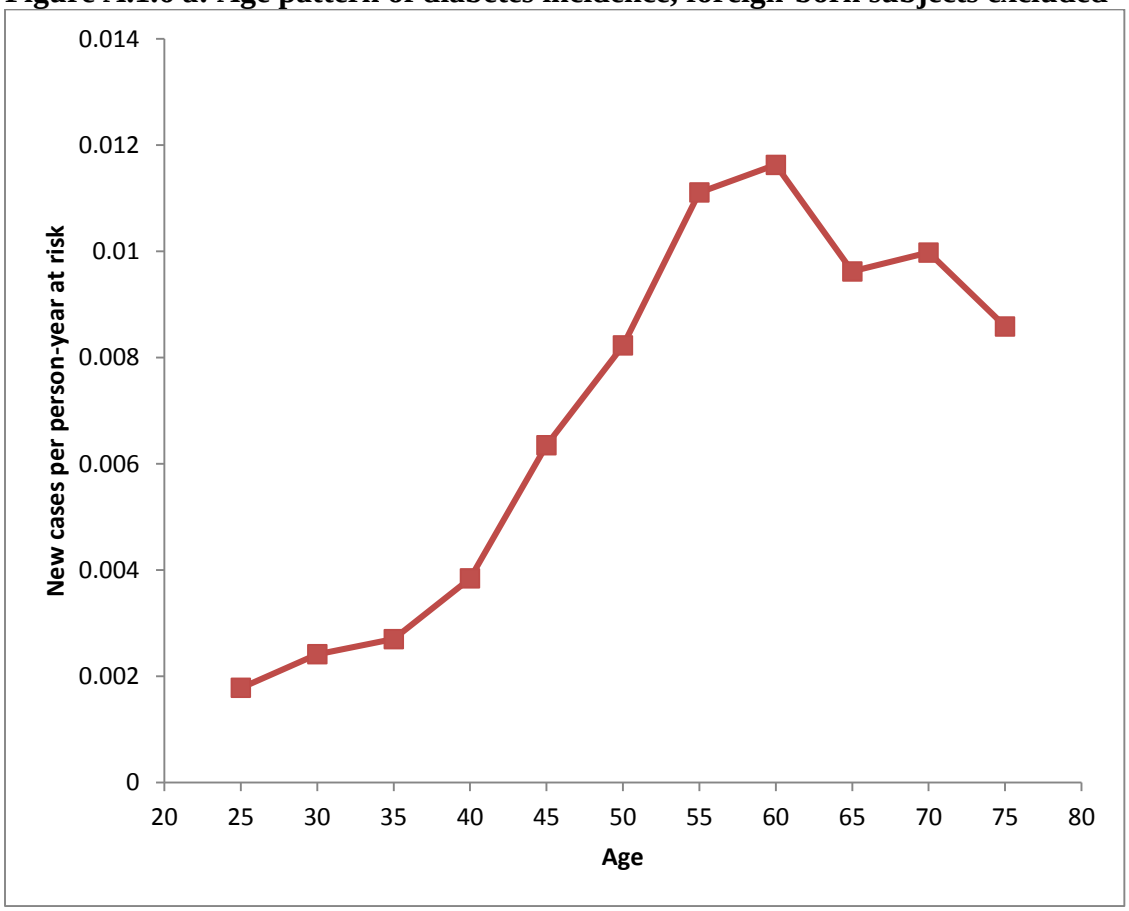
**Table A.1.5 a: Estimates of Diabetes Incidence for 1950-59 birth cohort**

| Age Interval | Incidence | Incidence with no differential mortality |
|---|---|---|
| 20-24 to 25-29 | 0.000994 | 0.000934 |
| 25-29 to 30-34 | 0.001693 | 0.001617 |
| 30-34 to 35-39 | 0.002422 | 0.002303 |
| 35-39 to 40-44 | 0.003611 | 0.003429 |
| 40-44 to 45-49 | 0.006399 | 0.00612 |
| 45-49 to 50-54 | 0.007944 | 0.007516 |
| 50-54 to 55-59 | 0.011015 | 0.010359 |
| 55-59 to 60-64 | 0.011268 | 0.010269 |
| 60-64 to 65-69 | 0.010361 | 0.008846 |
| 65-69 to 70-74 | 0.009888 | 0.007609 |
| 70-74 to 75-79 | 0.008723 | 0.006022 |

Incidence estimates are based on cohort prevalence estimates from age-cohort model and life-table values by diabetes status (nondiabetic versus entire population); see Methods section in text for details. Figure 1-6 in the main text plots the values in the "Incidence" column above. To demonstrate the effect of using mortality differences by diabetes status on the estimates, the table shows estimates of incidence that would result if we had ignored mortality differences by diabetes status.

### A.1.6. Incidence with foreign-born excluded from sample

**Figure A.1.6 a: Age pattern of diabetes incidence, foreign-born subjects excluded**




This calculation plots the 1950-59 birth cohort, but the shape of the curve is the same for all decadal birth cohorts. Foreign-born individuals were excluded from the sample. Data are from the National Health and Nutrition Examination Surveys (NHANES), United States, 1988 to 2010.

### A.1.7. Results based on "High Risk" of Diabetes, and Discussion of Threshold Choice

Appendix Figures A.1.7a and A.1.7b show estimates of the prevalence of "at least high risk" of diabetes, using HbA1c ≥ 6.0% (42 mmol/mol) (American Diabetes Association 2012), by period and cohort. Appendix Figure A.1.7c shows the cohort coefficients from the age/cohort model. Appendix Figure A.1.7d shows the period effects in the age/period and age/period/cohort models discussed in the Statistical Methods section, as applied to the threshold HbA1c ≥ 6.0%. Appendix Figure A.1.7e shows the modeled age-pattern of "at least high risk."

Although the recent ADA guidelines mention 6.0% as a possible threshold, they note that there is a "continuum of risk for diabetes with all glycemic measures" and did not formally identify 6.0% as a formal "high risk" threshold (American Diabetes Association 2012). A recent meta-analysis indicated that there is no clear HbA1c-based threshold above which the risk of incident diabetes increases dramatically (Gregg, Geiss, et al. 2013). Nevertheless, using the 6.0% threshold is a useful way to test the sensitivity of our methods to the choice of threshold. The patterns found using the 6.0% threshold are similar to the patterns found using the 6.5% threshold.

**Figure A.1.7 a: Prevalence of HbA1c 6.0%+ in Successive NHANES Waves**

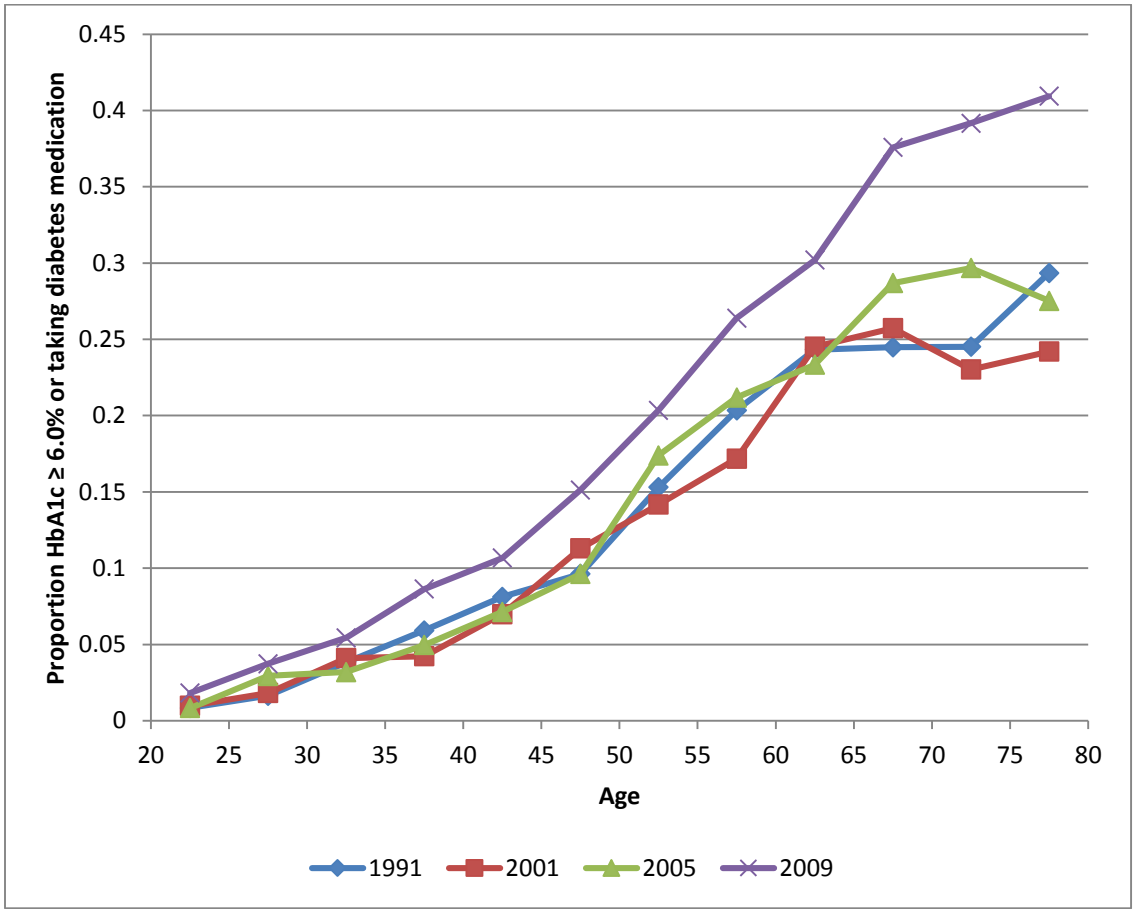


**Figure A.1.7 b: Prevalence of HbA1c 6.0%+ in Successive Decadal Birth Cohorts**

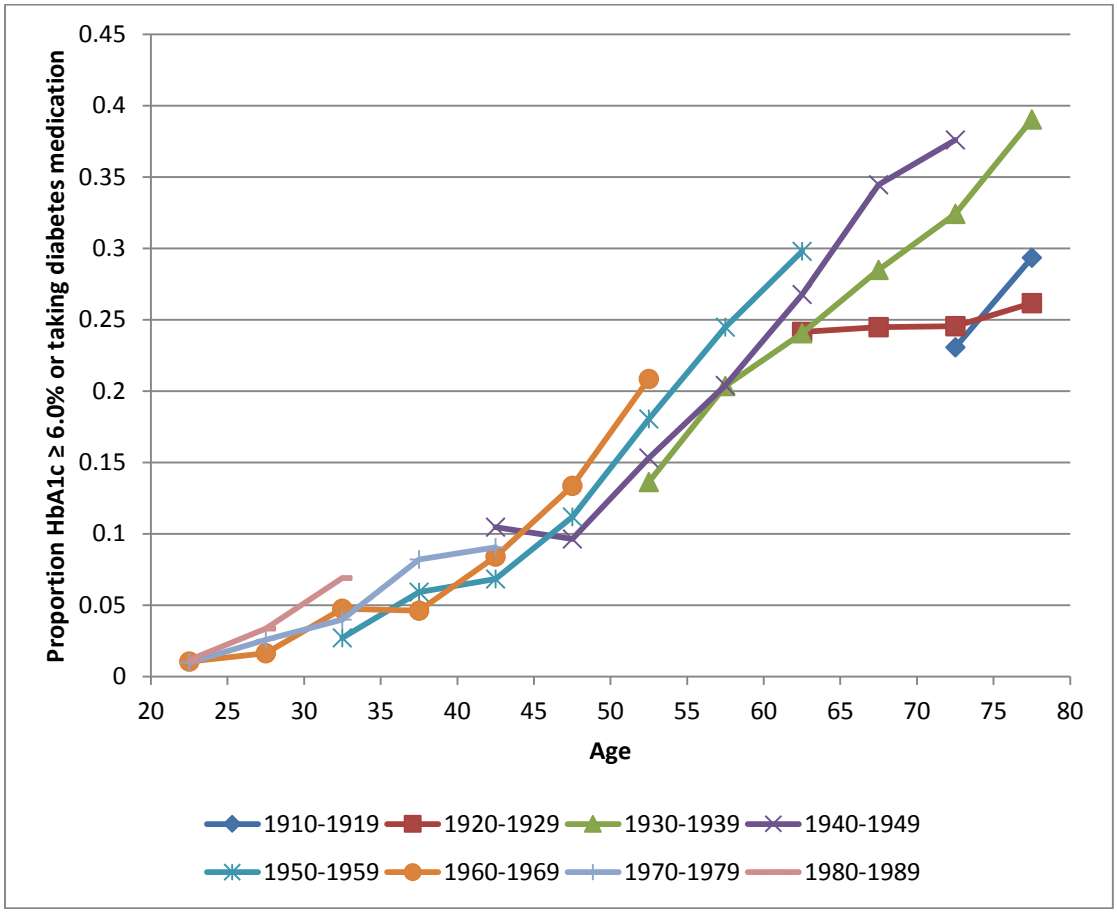


**Figure A.1.7 c: Age-Adjusted Prevalence of HbA1c 6.0%+ in Birth Cohorts Relative to 1910-1919 Birth Cohort**

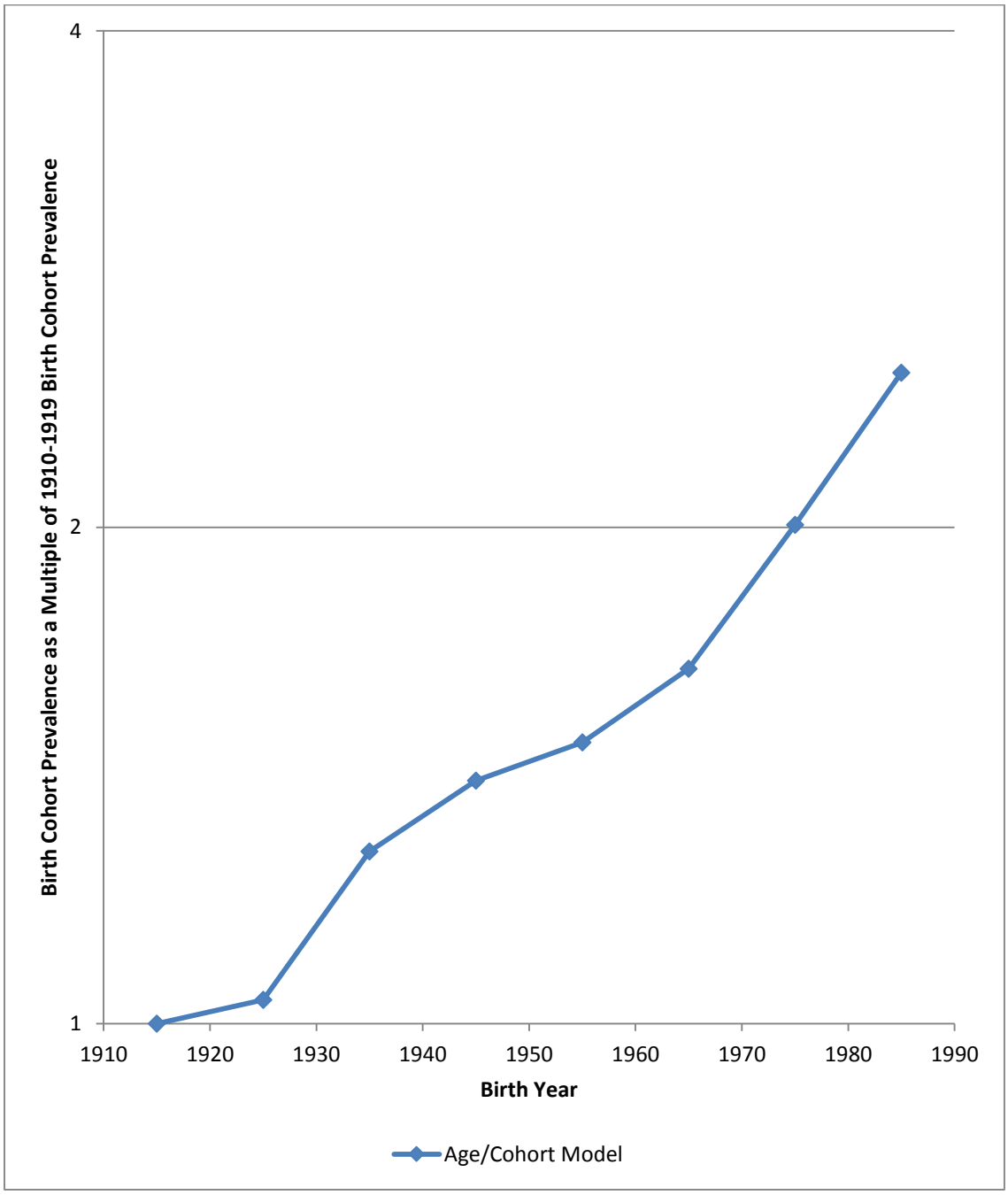


**Figure A.1.7 d: Prevalence of HbA1c 6.0%+ Relative to that in 1988-1994 Observation Period**

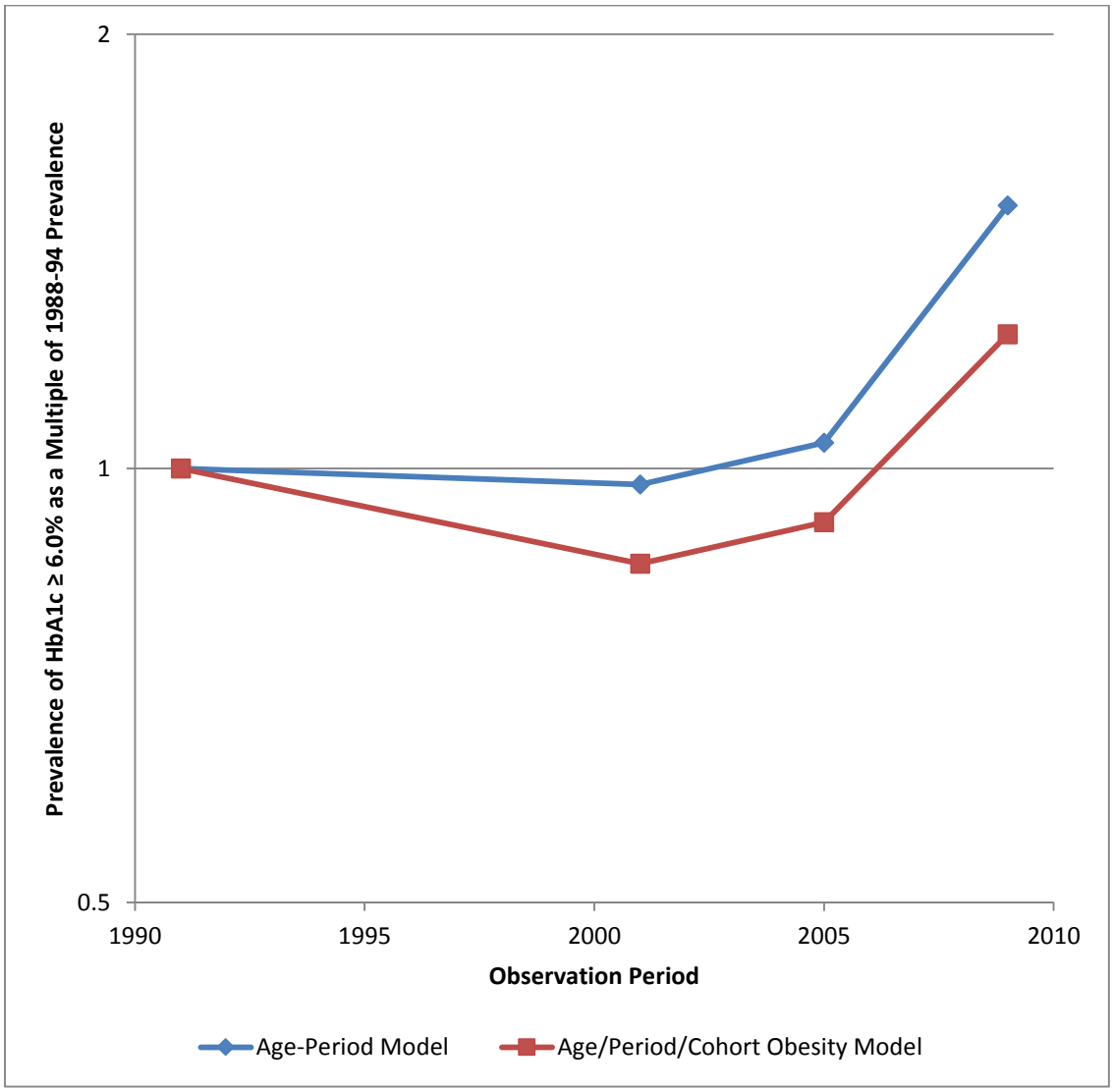


**Figure A.1.7 e: Age-Specific Prevalence of HbA1c 6.0%+ Relative to Age 20-24 Prevalence**

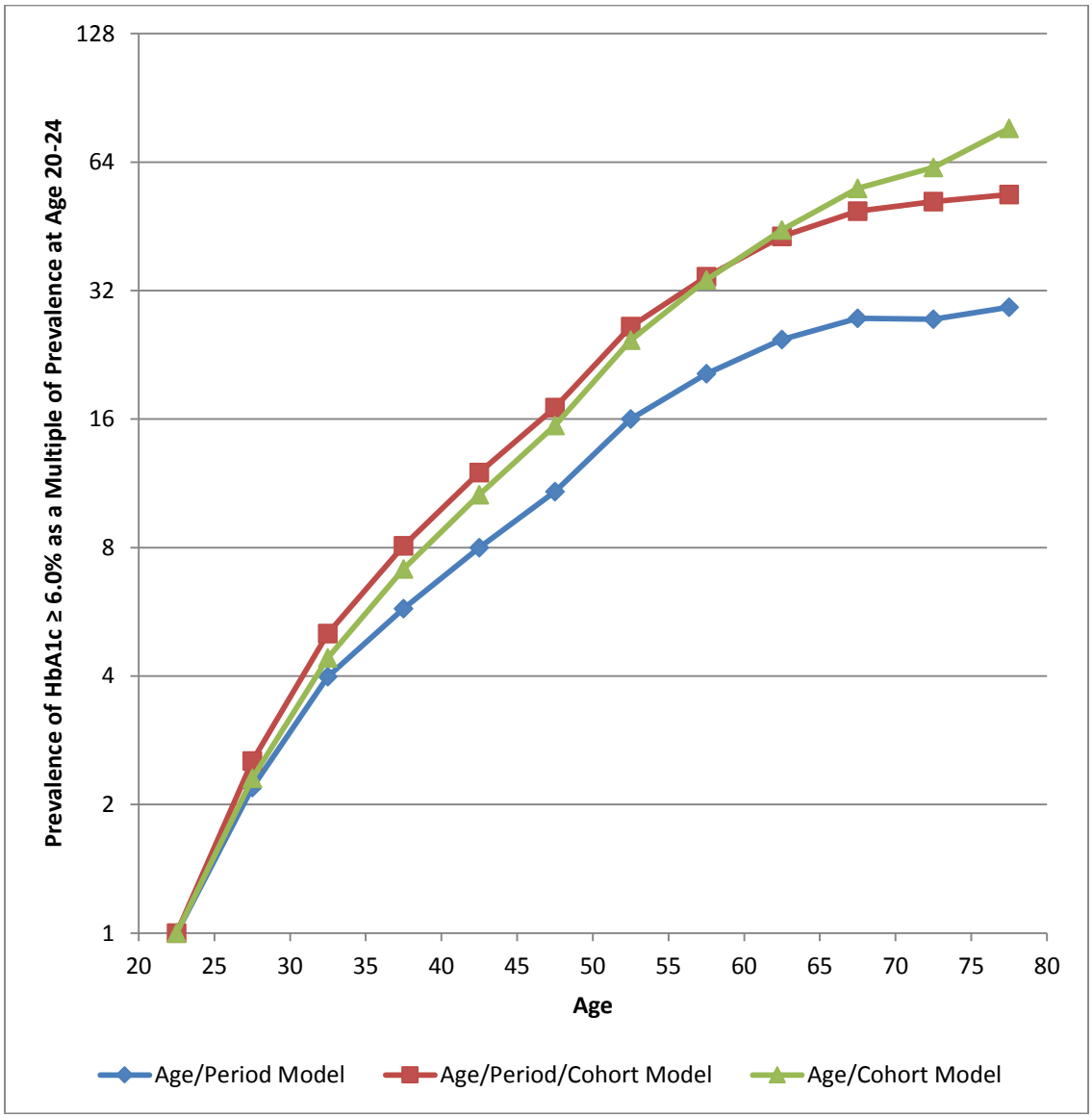


**Table A.1.7 a: Results of Age-Cohort Model**

Dependent variable = Log Prevalence of (HbA1c ≥ 6.0%)

| Indicator | Coefficient | SE | t-statistic | p-value | 95% CI | |
|---|---|---|---|---|---|---|
| Age 25-29 | 0.8303195 | 0.121278 | 6.846402 | 7.07E-07 | 0.578804 | 1.081835 |
| Age 30-34 | 1.474147 | 0.130009 | 11.33883 | 1.17E-10 | 1.204525 | 1.743768 |
| Age 35-39 | 1.944194 | 0.136118 | 14.28316 | 1.31E-12 | 1.661903 | 2.226485 |
| Age 40-44 | 2.33351 | 0.141649 | 16.47384 | 7.35E-14 | 2.039747 | 2.627273 |
| Age 45-49 | 2.691181 | 0.152291 | 17.67128 | 1.74E-14 | 2.375348 | 3.007013 |
| Age 50-54 | 3.129008 | 0.163882 | 19.09303 | 3.51E-15 | 2.789137 | 3.468879 |
| Age 55-59 | 3.4235 | 0.178286 | 19.20231 | 3.11E-15 | 3.053757 | 3.793242 |
| Age 60-64 | 3.654953 | 0.180346 | 20.26636 | 1.01E-15 | 3.280939 | 4.028967 |
| Age 65-69 | 3.828174 | 0.193689 | 19.76451 | 1.71E-15 | 3.426487 | 4.229861 |
| Age 70-74 | 3.876574 | 0.204934 | 18.91618 | 4.26E-15 | 3.451566 | 4.301582 |
| Age 75-79 | 4.00686 | 0.222296 | 18.02492 | 1.16E-14 | 3.545847 | 4.467873 |
| 1920-29 cohort | 0.0330163 | 0.184893 | 0.17857 | 0.85991 | -0.35043 | 0.41646 |
| 1930-39 cohort | 0.240484 | 0.18356 | 1.31011 | 0.203674 | -0.1402 | 0.621165 |
| 1940-49 cohort | 0.3391446 | 0.203955 | 1.66284 | 0.110527 | -0.08383 | 0.762121 |
| 1950-59 cohort | 0.3927891 | 0.218429 | 1.79825 | 0.085875 | -0.0602 | 0.845782 |
| 1960-69 cohort | 0.4955284 | 0.2296 | 2.158226 | 0.042089 | 0.019368 | 0.971689 |
| 1970-79 cohort | 0.6965528 | 0.241427 | 2.885149 | 0.008593 | 0.195864 | 1.197242 |
| 1980-89 cohort | 0.9089721 | 0.260882 | 3.48423 | 0.002102 | 0.367937 | 1.450008 |
| constant | -5.280235 | 0.252282 | -20.9299 | 5.14E-16 | -5.80344 | -4.75703 |
| R-Squared | 0.9878688 | | | | | |
| N | 41 | | | | | |

**Table A.1.7 b: Results of Age-Period Model**

Dependent variable = Log Prevalence of (HbA1c ≥ 6.0%)

| Indicator | Coefficient | SE | t-statistic | p-value | 95% CI | |
|---|---|---|---|---|---|---|
| Age 25-29 | 0.785042 | 0.089818 | 8.740355 | 4.21E-10 | 0.602306 | 0.967778 |
| Age 30-34 | 1.382502 | 0.089036 | 15.52754 | 9.9E-17 | 1.201358 | 1.563646 |
| Age 35-39 | 1.750047 | 0.088906 | 19.68429 | 8.32E-20 | 1.569167 | 1.930927 |
| Age 40-44 | 2.079389 | 0.087776 | 23.6898 | 2.73E-22 | 1.900808 | 2.25797 |
| Age 45-49 | 2.381469 | 0.091704 | 25.96907 | 1.53E-23 | 2.194896 | 2.568043 |
| Age 50-54 | 2.77329 | 0.092897 | 29.85342 | 1.85E-25 | 2.58429 | 2.962291 |
| Age 55-59 | 3.016927 | 0.098845 | 30.52183 | 9.13E-26 | 2.815825 | 3.218028 |
| Age 60-64 | 3.201748 | 0.089644 | 35.71608 | 5.92E-28 | 3.019365 | 3.384131 |
| Age 65-69 | 3.317046 | 0.094479 | 35.10868 | 1.03E-27 | 3.124826 | 3.509265 |
| Age 70-74 | 3.311332 | 0.095643 | 34.6218 | 1.61E-27 | 3.116745 | 3.505919 |
| Age 75-79 | 3.376998 | 0.10645 | 31.7239 | 2.66E-26 | 3.160425 | 3.593572 |
| 2001 NHANES | -0.0258708 | 0.054771 | -0.47235 | 0.639789 | -0.1373 | 0.085562 |
| 2005 NHANES | 0.0411817 | 0.054802 | 0.751463 | 0.4577 | -0.07031 | 0.152677 |
| 2009 NHANES | 0.4200199 | 0.050525 | 8.313196 | 1.33E-09 | 0.317227 | 0.522813 |
| constant | -4.683064 | 0.067184 | -69.7048 | 2.09E-37 | -4.81975 | -4.54638 |
| R-Squared | 0.9897704 | | | | | |
| N | 48 | | | | | |

**Table A.1.7 c: results of Age-Period-Cohort Obesity Model**

Dependent variable = Log Prevalence of (HbA1c ≥ 6.0%)

| Indicator | Coefficient | SE | t-statistic | p-value | 95% CI | |
|---|---|---|---|---|---|---|
| Age 25-29 | 0.9286565 | 0.087923 | 10.5622 | 1.88E-11 | 0.748835 | 1.108479 |
| Age 30-34 | 1.61597 | 0.102132 | 15.82242 | 8.41E-16 | 1.407087 | 1.824853 |
| Age 35-39 | 2.088447 | 0.12797 | 16.31986 | 3.74E-16 | 1.82672 | 2.350175 |
| Age 40-44 | 2.484058 | 0.146681 | 16.93516 | 1.41E-16 | 2.184062 | 2.784053 |
| Age 45-49 | 2.834346 | 0.16289 | 17.40037 | 6.89E-17 | 2.501198 | 3.167493 |
| Age 50-54 | 3.271163 | 0.177394 | 18.44012 | 1.46E-17 | 2.908352 | 3.633973 |
| Age 55-59 | 3.541548 | 0.188307 | 18.8073 | 8.63E-18 | 3.156417 | 3.92668 |
| Age 60-64 | 3.757913 | 0.195096 | 19.26188 | 4.54E-18 | 3.358897 | 4.156929 |
| Age 65-69 | 3.893697 | 0.203847 | 19.1011 | 5.69E-18 | 3.476784 | 4.31061 |
| Age 70-74 | 3.946568 | 0.221865 | 17.78819 | 3.83E-17 | 3.492804 | 4.400332 |
| Age 75-79 | 3.984019 | 0.235082 | 16.94739 | 1.39E-16 | 3.503223 | 4.464814 |
| 2001 NHANES | -0.1520016 | 0.06738 | -2.25588 | 0.031793 | -0.28981 | -0.01419 |
| 2005 NHANES | -0.0857865 | 0.067733 | -1.26654 | 0.2154 | -0.22432 | 0.052743 |
| 2009 NHANES | 0.2139795 | 0.080386 | 2.661893 | 0.012539 | 0.049571 | 0.378388 |
| obesity | 4.388277 | 1.597306 | 2.747298 | 0.010222 | 1.121419 | 7.655136 |
| constant | -5.274105 | 0.202519 | -26.0426 | 1.15E-21 | -5.6883 | -4.85991 |
| R-Squared | 0.9925176 | | | | | |
| N | 45 | | | | | |

## A.1.8. ALTERNATIVE DELINEATION OF BIRTH COHORTS

**Figure A.1.8 a: Age-specific diabetes prevalence in successive ten-year birth cohorts, using alternative delineation of birth cohorts**

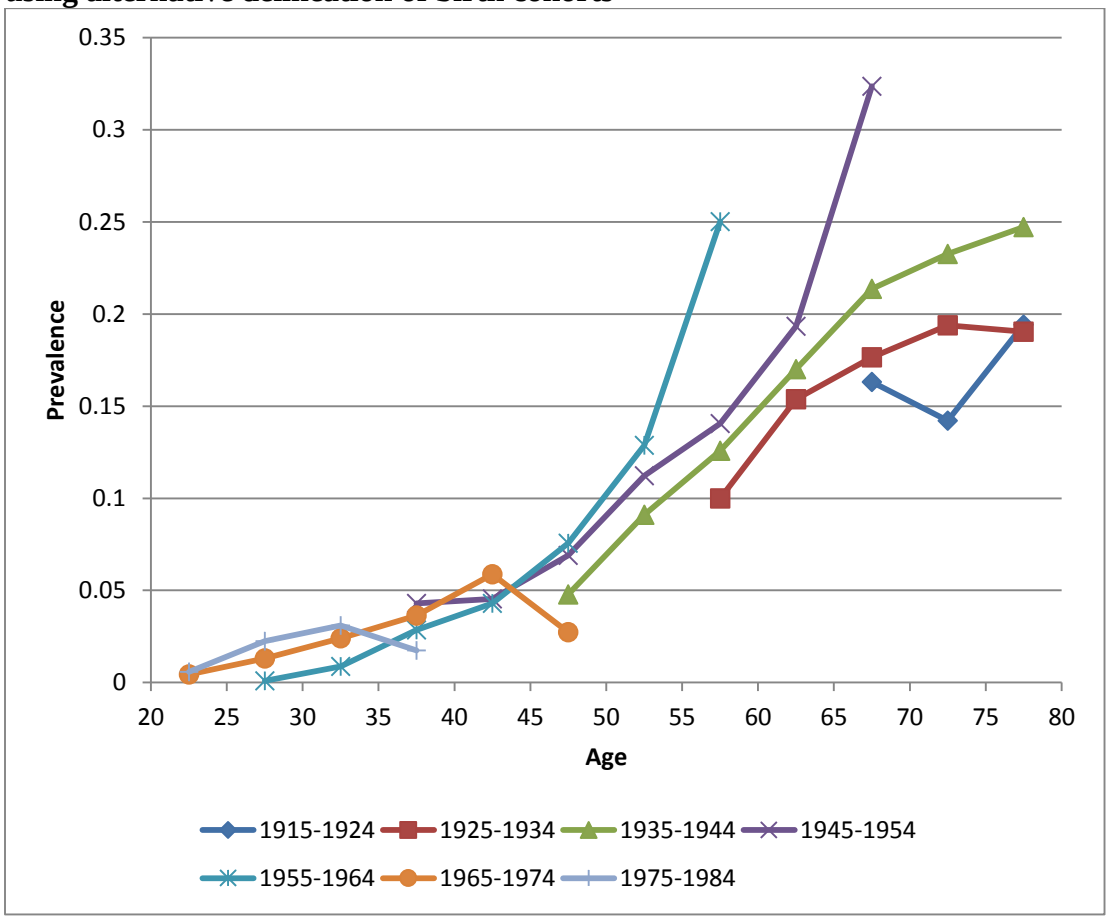
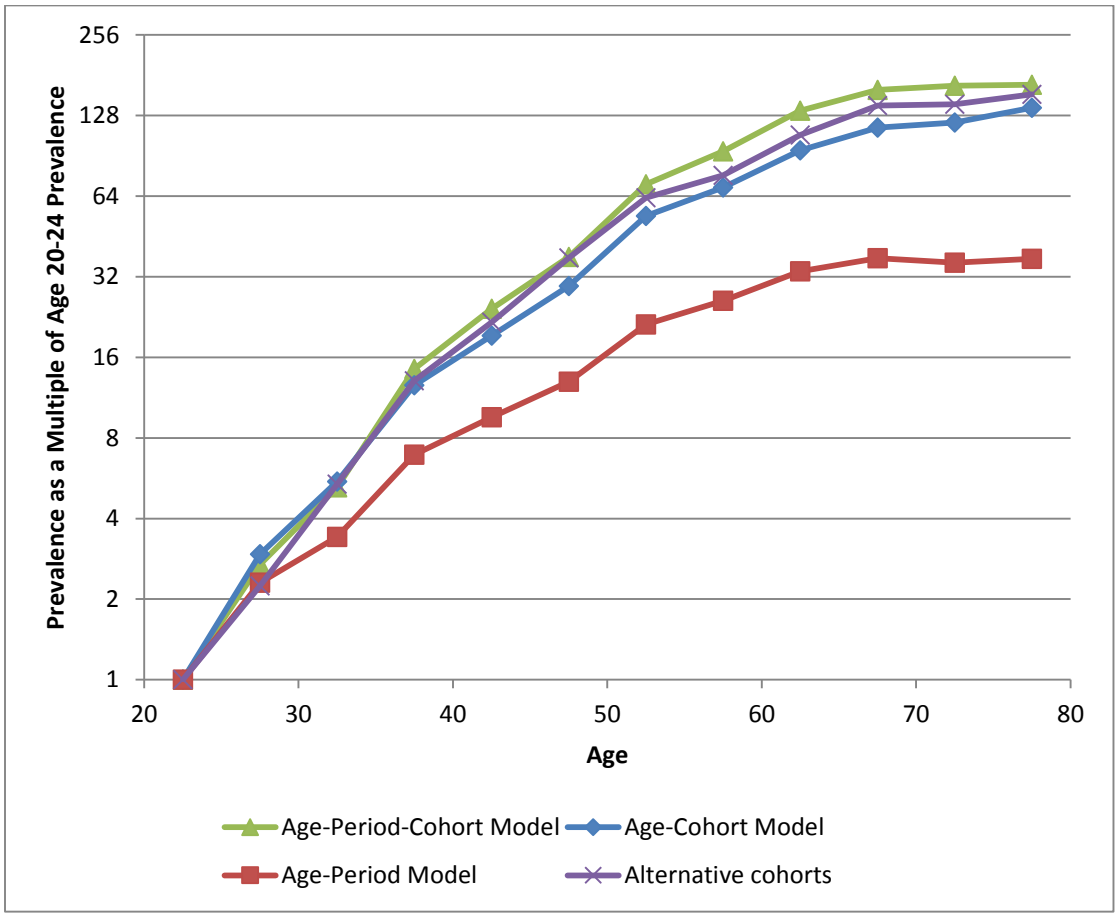


**Figure A.1.8 b: Age pattern of diabetes prevalence, including alternative delineation of birth cohorts**




The "alternative cohorts" are born 1915-1924, 1925-1934, etc., rather than 1910-1919, 1920-1929, etc.

Data are from the National Health and Nutrition Examination Surveys (NHANES), United States, 1988 to 2010.

## A.2. Appendices to Chapter 2 (Lifetime Risk of Dementia)

### A.2.1. Models of dementia incidence, prevalence, and differential mortality

A) Prevalence model: $logit(prevalence_x) = \alpha + \beta x$,

Point estimates:

| Term | Coefficient estimate |
|---|---|
| Age ($x$) | 0.152565 |
| Constant | -14.2737 |

Variance-covariance:

| | Age | Constant |
|---|---|---|
| Age | 0.000295 | -0.02424 |
| Constant | -0.02424 | 2.004415 |

B) Incidence model: $logit(incidence_x) = \alpha' + \beta' x$,

Point estimates:

| Term | Coefficient estimate |
|---|---|
| Age ($x$) | 0.087151 |
| Constant | -10.6868 |

Variance-covariance:

| | Age | Constant |
|---|---|---|
| Age | 0.000407 | -0.03363 |
| Constant | -0.03363 | 2.793747 |

C) Differential mortality

Model: $\ln(m_{x,dem}) = \alpha + \beta_1 x + \beta_2 dementia + \beta_3 x * dementia.$

Point estimates:

| Term | Coefficient estimate |
|---|---|
| Dementia | 6.435545 |
| Age ($x$) | 0.110955 |
| Age*Dementia | -0.06139 |
| Constant | -12.2631 |

Variance-covariance:

| | Dementia | Age | Age*Dementia | Constant |
|---|---|---|---|---|
| Dementia | 3.427942 | 0.027964 | -0.03975 | -2.35307 |
| Age | 0.027964 | 0.000299 | -0.00033 | -0.025 |
| Age*Dementia | -0.03975 | -0.00033 | 0.000464 | 0.027693 |
| Constant | -2.35307 | -0.025 | 0.027693 | 2.10715 |

### A.2.2. Deterministic inputs and results

**Table A.2.2 a: Deterministic inputs to stationary approach**

| Age | Estimated Prevalence | Estimated Incidence | Implied RR |
|---|---|---|---|
| 70 | 0.029 | 0.010 | 10.213 |
| 75 | 0.060 | 0.016 | 4.041 |
| 80 | 0.120 | 0.024 | 1.824 |
| 85 | 0.226 | 0.036 | 1.062 |
| 90 | 0.385 | 0.055 | 0.955 |
| 95 | 0.574 | 0.083 | 0.962 |
| 100 | 0.743 | 0.122 | 1.150 |

RR = Mortality rate ratio, with dementia vs. without dementia.

Parametric models were fit to ADAMS baseline data for prevalence, ADAMS longitudinal data for incidence. Parameters of the models – shown in Appendix Table 1 – were treated as non-stochastic and run through life table operations as described in the Methods section under Approach 1.

Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For prevalence, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1920 birth cohort.

**Table A.2.2 b: Deterministic life cycle quantities using stationary approach**

A. Males

| Age | LE | DFLE | DLE | Lifetime Risk | DFLE' |
|---|---|---|---|---|---|
| 70 | 12.31 | 10.84 | 1.47 | 0.233 | 11.14 |
| 75 | 9.65 | 8.05 | 1.60 | 0.230 | 8.52 |
| 80 | 7.26 | 5.54 | 1.72 | 0.222 | 6.24 |
| 85 | 5.20 | 3.43 | 1.77 | 0.209 | 4.36 |
| 90 | 3.64 | 1.90 | 1.75 | 0.199 | 3.01 |
| 95 | 2.61 | 0.95 | 1.66 | 0.200 | 2.14 |
| 100 | 2.02 | 0.45 | 1.57 | 0.228 | 1.71 |

B. Females

| Age | LE | DFLE | DLE | Lifetime Risk | DFLE' |
|---|---|---|---|---|---|
| 70 | 15.25 | 13.04 | 2.21 | 0.312 | 13.40 |
| 75 | 11.91 | 9.62 | 2.29 | 0.298 | 10.18 |
| 80 | 8.91 | 6.57 | 2.34 | 0.280 | 7.40 |
| 85 | 6.37 | 4.06 | 2.32 | 0.259 | 5.16 |
| 90 | 4.42 | 2.22 | 2.19 | 0.241 | 3.52 |
| 95 | 3.10 | 1.09 | 2.00 | 0.236 | 2.46 |
| 100 | 2.32 | 0.50 | 1.81 | 0.255 | 1.89 |

LE = total life expectancy for a randomly chosen person in the population of given age. DFLE = Dementia-free life expectancy for a randomly chosen person in the population of given age, DLE = life expectancy with dementia for a randomly chosen person in the population of given age, Lifetime Risk = probability that a dementia-free person will develop dementia later in life. DFLE' = Dementia-free life expectancy for a dementia-free person in the population of given age.

Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For prevalence, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1920 birth cohort. Prevalence, incidence, and mortality rate ratios used in this table are shown in Appendix Table 2.

**Table A.2.2 c: Deterministic inputs to non-stationary approach**

| Age | Implied Prevalence | Estimated Incidence | Estimated RR |
|---|---|---|---|
| 70 | 0.029 | 0.010 | 8.486 |
| 75 | 0.057 | 0.016 | 6.243 |
| 80 | 0.090 | 0.024 | 4.593 |
| 85 | 0.134 | 0.036 | 3.379 |
| 90 | 0.191 | 0.055 | 2.486 |
| 95 | 0.277 | 0.083 | 1.829 |
| 100 | 0.443 | 0.122 | 1.346 |

Parametric models were fit to ADAMS longitudinal data for incidence and differential mortality. Parameters of the models – shown in Appendix Table 1 – were treated as non-stochastic and run through life table operations as described in the Methods section.

Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For relative risk of death, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1920 birth cohort. Mortality by dementia status data from Health and Retirement Study (HRS), United States, 2001-2011.

**Table A.2.2 d: Deterministic life cycle quantities using non-stationary approach**

A. Males

| Age | LE | DFLE | DLE | Lifetime Risk | DFLE' |
|---|---|---|---|---|---|
| 70 | 12.31 | 11.58 | 0.73 | 0.267 | 11.93 |
| 75 | 9.65 | 8.93 | 0.72 | 0.269 | 9.37 |
| 80 | 7.26 | 6.53 | 0.73 | 0.264 | 7.05 |
| 85 | 5.20 | 4.49 | 0.71 | 0.253 | 5.05 |
| 90 | 3.64 | 2.95 | 0.69 | 0.239 | 3.51 |
| 95 | 2.61 | 1.87 | 0.73 | 0.232 | 2.44 |
| 100 | 2.02 | 1.10 | 0.91 | 0.234 | 1.76 |

B. Females

| Age | LE | DFLE | DLE | Lifetime Risk | DFLE' |
|---|---|---|---|---|---|
| 70 | 15.25 | 13.55 | 1.70 | 0.344 | 13.95 |
| 75 | 11.91 | 10.25 | 1.66 | 0.338 | 10.90 |
| 80 | 8.91 | 7.38 | 1.53 | 0.325 | 8.17 |
| 85 | 6.37 | 5.02 | 1.35 | 0.307 | 5.86 |
| 90 | 4.42 | 3.24 | 1.18 | 0.285 | 4.04 |
| 95 | 3.10 | 2.00 | 1.09 | 0.269 | 2.77 |
| 100 | 2.32 | 1.13 | 1.18 | 0.264 | 1.95 |

LE = total life expectancy for a randomly chosen person in the population of given age. DFLE = Dementia-free life expectancy for a randomly chosen person in the population of given age, DLE = life expectancy with dementia for a randomly chosen person in the population of given age, Lifetime Risk = probability that a dementia-free person will develop dementia later in life. DFLE' = Dementia-free life expectancy for a dementia-free person in the population of given age.

Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For relative risk of death, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1920 birth cohort. Mortality by dementia status data from Health and Retirement Study (HRS), United States, 2001-2011. Prevalence, incidence, and mortality rate ratios used in this table are shown in Appendix Table 4.

**Table A.2.2 e: Deterministic life cycle quantities using 1940 cohort life table, non-stationary approach**

A. Males

| Age | LE | DFLE | DLE | Lifetime Risk | DFLE' |
|---|---|---|---|---|---|
| 70 | 13.64 | 12.58 | 1.06 | 0.305 | 12.95 |
| 75 | 10.65 | 9.60 | 1.04 | 0.303 | 10.15 |
| 80 | 7.96 | 6.96 | 1.00 | 0.295 | 7.61 |
| 85 | 5.70 | 4.77 | 0.93 | 0.280 | 5.46 |
| 90 | 4.05 | 3.16 | 0.90 | 0.268 | 3.83 |
| 95 | 2.95 | 2.00 | 0.95 | 0.260 | 2.69 |
| 100 | 2.30 | 1.16 | 1.14 | 0.262 | 1.94 |

B. Females

| Age | LE | DFLE | DLE | Lifetime Risk | DFLE' |
|---|---|---|---|---|---|
| 70 | 15.99 | 14.05 | 1.93 | 0.371 | 14.47 |
| 75 | 12.62 | 10.71 | 1.91 | 0.366 | 11.41 |
| 80 | 9.57 | 7.76 | 1.81 | 0.355 | 8.64 |
| 85 | 6.93 | 5.30 | 1.63 | 0.336 | 6.26 |
| 90 | 4.89 | 3.43 | 1.45 | 0.316 | 4.38 |
| 95 | 3.49 | 2.11 | 1.37 | 0.300 | 3.03 |
| 100 | 2.63 | 1.18 | 1.45 | 0.294 | 2.15 |

Notes: This table uses as inputs the quantities shown in Appendix Table 4, along with 1940 birth cohort life tables from the U.S. Social Security Administration.

LE = total life expectancy for a randomly chosen person in the population of given age. DFLE = Dementia-free life expectancy for a randomly chosen person in the population of given age, DLE = life expectancy with dementia for a randomly chosen person in the population of given age, Lifetime Risk = probability that a dementia-free person will develop dementia later in life.

Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For relative risk of death, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1940 birth cohort. Mortality by dementia status data from Health and Retirement Study (HRS), United States, 2001-2011. Prevalence, incidence, and mortality rate ratios used in this table are shown in Appendix Table 4.

### A.2.3. Incidence rates under intervention scenarios

*New cases per dementia-free person-year lived*

| Age | Scenario 1&2 | Scenario 3&4 | Scenario 5 |
|---|---|---|---|
| 70 | 0.0095 | 0.00685 | 0.0058 |
| 75 | 0.0144 | 0.01032 | 0.0084 |
| 80 | 0.0220 | 0.01566 | 0.0122 |
| 85 | 0.0337 | 0.02394 | 0.0178 |
| 90 | 0.0518 | 0.03675 | 0.0261 |
| 95 | 0.0794 | 0.05648 | 0.0384 |
| 100 | 0.1201 | 0.08636 | 0.0565 |

Scenarios 1 & 2 delay dementia incidence by one year compared to rates estimated from ADAMS and shown in Table 1. Scenarios 3 & 4 delay dementia incidence by five years. Scenario 5 reduces the acceleration of dementia incidence with age by 10%.

### A.2.4. Approximated dementia-free survivors to middle of age interval versus dementia-free person-years lived in the interval

Males, non-stationary approach, deterministic calculation

| Age | $\hat{L}_x^{ND}$ | $L_x^{ND}$ | Absolute Difference | Diff. as % of $L_x^{ND}$ |
|---|---|---|---|---|
| 71 | 91351.4 | 91329.4 | 22.01 | 0.02% |
| 72 | 87440.3 | 87376.9 | 63.41 | 0.07% |
| 73 | 83361.0 | 83292.3 | 68.68 | 0.08% |
| 74 | 79139.0 | 79140.9 | 1.89 | 0.00% |
| 75 | 74991.4 | 74947.0 | 44.39 | 0.06% |
| 76 | 70708.7 | 70676.5 | 32.14 | 0.05% |
| 77 | 66374.0 | 66352.5 | 21.43 | 0.03% |
| 78 | 62007.1 | 61983.9 | 23.16 | 0.04% |
| 79 | 57592.1 | 57561.6 | 30.58 | 0.05% |
| 80 | 53108.6 | 53124.7 | 16.08 | 0.03% |
| 81 | 48703.9 | 48719.4 | 15.48 | 0.03% |
| 82 | 44329.6 | 44306.6 | 23.00 | 0.05% |
| 83 | 39870.7 | 39899.1 | 28.40 | 0.07% |
| 84 | 35520.1 | 35556.9 | 36.79 | 0.10% |
| 85 | 31251.5 | 31309.9 | 58.48 | 0.19% |
| 86 | 27121.4 | 27202.5 | 81.03 | 0.30% |
| 87 | 23176.1 | 23282.8 | 106.73 | 0.46% |
| 88 | 19469.8 | 19599.9 | 130.06 | 0.66% |
| 89 | 16047.1 | 16200.4 | 153.31 | 0.95% |
| 90 | 12954.2 | 13123.7 | 169.55 | 1.29% |
| 91 | 10216.6 | 10399.1 | 182.48 | 1.75% |
| 92 | 7857.0 | 8044.1 | 187.15 | 2.33% |
| 93 | 5876.3 | 6061.0 | 184.73 | 3.05% |
| 94 | 4262.6 | 4437.8 | 175.24 | 3.95% |
| 95 | 2990.0 | 3152.1 | 162.16 | 5.14% |
| 96 | 2028.6 | 2169.3 | 140.71 | 6.49% |
| 97 | 1327.1 | 1444.4 | 117.28 | 8.12% |
| 98 | 836.8 | 930.6 | 93.80 | 10.08% |
| 99 | 510.6 | 579.7 | 69.12 | 11.92% |
| 100 | 298.0 | 348.5 | 50.55 | 14.51% |

The $\hat{L}_x^{ND}$ values are approximated dementia-free survivors; $L_x^{ND}$ are person-years lived without dementia in the age interval. The $\hat{L}_x^{ND}$ values were used to estimate the dementia prevalence in each age group using the non-stationary approach and were estimated assuming linearity of survival within one-year age intervals; see Methods section, under Approach 2, for details.

### A.2.5. Results using alternative censoring strategy

**Table A.2.5 a: Incidence, Prevalence, and Differential Mortality, censoring subjects without dementia at end of ADAMS study period (Stationary Approach)**

| | Fitted prevalence & incidence | | | | Implied | |
|---|---|---|---|---|---|---|
| Age | Prevalence | (SE) | Incidence | (SE) | RR | (SE) |
| 70 | 0.030 | (0.007) | 0.010 | (0.003) | 13.542 | (13.668) |
| 75 | 0.061 | (0.009) | 0.015 | (0.003) | 3.939 | (1.921) |
| 80 | 0.121 | (0.012) | 0.022 | (0.003) | 1.487 | (0.101) |
| 85 | 0.227 | (0.022) | 0.032 | (0.005) | 0.889 | (0.261) |
| 90 | 0.384 | (0.043) | 0.046 | (0.009) | 0.875 | (0.236) |
| 95 | 0.569 | (0.061) | 0.068 | (0.019) | 0.919 | (0.202) |
| 100 | 0.734 | (0.064) | 0.099 | (0.035) | 1.106 | (0.228) |

RR = Relative risk of death, with dementia vs. without dementia

Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For prevalence, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1920 birth cohort.

Parametric models were fitted to prevalence and incidence data to generate single-year age-specific estimates. Subjects who survived ADAMS without a diagnosis of dementia were censored at the end of the ADAMS study period, rather than at their last assessment. RR was implied by fitted prevalence, incidence, and stationary-population relations.

**Table A.2.5 b: Life cycle quantities for dementia, stationary approach, censoring subjects without dementia at end of ADAMS study period**

A) Males

| Age | LE | DFLE | DLE | (SE) | Lifetime Risk | (SE) | DFLE' | (SE) |
|---|---|---|---|---|---|---|---|---|
| 70 | 12.31 | 10.83 | 1.48 | (0.129) | 0.214 | (0.028) | 11.14 | (0.119) |
| 75 | 9.65 | 8.04 | 1.61 | (0.142) | 0.206 | (0.025) | 8.53 | (0.134) |
| 80 | 7.26 | 5.54 | 1.72 | (0.162) | 0.194 | (0.028) | 6.25 | (0.143) |
| 85 | 5.20 | 3.43 | 1.76 | (0.182) | 0.179 | (0.035) | 4.37 | (0.135) |
| 90 | 3.64 | 1.91 | 1.74 | (0.187) | 0.168 | (0.044) | 3.01 | (0.109) |
| 95 | 2.61 | 0.96 | 1.64 | (0.163) | 0.166 | (0.056) | 2.15 | (0.078) |
| 100 | 2.02 | 0.47 | 1.55 | (0.123) | 0.187 | (0.073) | 1.71 | (0.039) |

B) Females

| Age | LE | DFLE | DLE | (SE) | Lifetime Risk | (SE) | DFLE' | (SE) |
|---|---|---|---|---|---|---|---|---|
| 70 | 15.25 | 13.17 | 2.08 | (0.190) | 0.283 | (0.034) | 13.41 | (0.191) |
| 75 | 11.91 | 9.77 | 2.14 | (0.205) | 0.265 | (0.033) | 10.19 | (0.205) |
| 80 | 8.91 | 6.74 | 2.16 | (0.227) | 0.244 | (0.038) | 7.41 | (0.209) |
| 85 | 6.37 | 4.25 | 2.12 | (0.249) | 0.221 | (0.047) | 5.17 | (0.190) |
| 90 | 4.42 | 2.41 | 2.00 | (0.251) | 0.203 | (0.056) | 3.53 | (0.149) |
| 95 | 3.10 | 1.26 | 1.83 | (0.220) | 0.196 | (0.067) | 2.47 | (0.102) |
| 100 | 2.32 | 0.63 | 1.69 | (0.167) | 0.210 | (0.085) | 1.90 | (0.061) |

LE = total life expectancy for a randomly chosen person in the population of given age. DFLE = Dementia-free life expectancy for a randomly chosen person in the population of given age, DLE = life expectancy with dementia for a randomly chosen person in the population of given age, Lifetime Risk = probability that a dementia-free person will develop dementia later in life, DFLE' = Dementia-free life expectancy for a dementia-free person of given age.
By construction, DFLE and DLE have the same standard error, and LE has zero variance.

Quantities were calculated using fitted values of dementia incidence and relative risk of death (with dementia vs. without) shown in Appendix Table 9.

Subjects who survived ADAMS without a diagnosis of dementia were censored at the end of the ADAMS study period, rather than at their last assessment.

Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For prevalence, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1920 birth cohort.

**Table A.2.5 c: Incidence, Prevalence, and Differential Mortality, censoring subjects without dementia at end of ADAMS study period (non-stationary approach)**

| Age | Implied Prevalence | (SE) | Estimated Incidence | (SE) | Estimated RR | (SE) |
|---|---|---|---|---|---|---|
| 70 | 0.030 | (0.173) | 0.010 | (0.003) | 10.56 | (4.290) |
| 75 | 0.052 | (0.228) | 0.015 | (0.003) | 7.94 | (2.322) |
| 80 | 0.075 | (0.274) | 0.022 | (0.003) | 6.04 | (1.200) |
| 85 | 0.098 | (0.313) | 0.032 | (0.005) | 4.67 | (0.699) |
| 90 | 0.119 | (0.345) | 0.046 | (0.009) | 3.65 | (0.664) |
| 95 | 0.145 | (0.381) | 0.068 | (0.019) | 2.89 | (0.777) |
| 100 | 0.206 | (0.454) | 0.099 | (0.035) | 2.33 | (0.874) |

RR = Relative risk of death, with dementia vs. without dementia

Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For relative risk of death, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1920 birth cohort. Mortality by dementia status data from Health and Retirement Study (HRS), United States, 2001-2011.

Parametric models were fitted to incidence and mortality data from ADAMS to generate single-year age-specific estimates. Prevalence in the cohort arises from the life table relations as described in the Methods section, under Approach 2.

Subjects who survived ADAMS without a diagnosis of dementia were censored at the end of the ADAMS study period, rather than at their last assessment.

**Table A.2.5 d: Life cycle quantities for dementia, non-stationary approach, censoring subjects without dementia at end of ADAMS study period**

A) Males

| Age | LE | DFLE | DLE | (SE) | Lifetime Risk | (SE) | DFLE' | (SE) |
|---|---|---|---|---|---|---|---|---|
| 70 | 12.31 | 11.83 | 0.48 | (0.186) | 0.252 | (0.033) | 12.19 | (0.150) |
| 75 | 9.65 | 9.21 | 0.44 | (0.182) | 0.249 | (0.033) | 9.63 | (0.106) |
| 80 | 7.26 | 6.85 | 0.42 | (0.163) | 0.240 | (0.036) | 7.28 | (0.069) |
| 85 | 5.20 | 4.82 | 0.37 | (0.140) | 0.225 | (0.044) | 5.24 | (0.042) |
| 90 | 3.64 | 3.31 | 0.34 | (0.119) | 0.211 | (0.054) | 3.66 | (0.035) |
| 95 | 2.61 | 2.28 | 0.33 | (0.112) | 0.205 | (0.065) | 2.59 | (0.036) |
| 100 | 2.02 | 1.63 | 0.39 | (0.136) | 0.216 | (0.083) | 1.95 | (0.047) |

B) Females

| Age | LE | DFLE | DLE | (SE) | Lifetime Risk | (SE) | DFLE' | (SE) |
|---|---|---|---|---|---|---|---|---|
| 70 | 15.25 | 13.84 | 1.41 | (0.249) | 0.323 | (0.038) | 14.27 | (0.234) |
| 75 | 11.91 | 10.59 | 1.32 | (0.255) | 0.312 | (0.039) | 11.22 | (0.157) |
| 80 | 8.91 | 7.76 | 1.14 | (0.245) | 0.295 | (0.044) | 8.46 | (0.111) |
| 85 | 6.37 | 5.45 | 0.92 | (0.220) | 0.274 | (0.052) | 6.10 | (0.071) |
| 90 | 4.42 | 3.70 | 0.72 | (0.186) | 0.252 | (0.062) | 4.24 | (0.042) |
| 95 | 3.10 | 2.51 | 0.59 | (0.160) | 0.238 | (0.073) | 2.95 | (0.035) |
| 100 | 2.32 | 1.74 | 0.57 | (0.164) | 0.242 | (0.090) | 2.18 | (0.044) |

LE = total life expectancy for a randomly chosen person in the population of given age. DFLE = Dementia-free life expectancy for a randomly chosen person in the population of given age, DLE = life expectancy with dementia for a randomly chosen person in the population of given age, Lifetime Risk = probability that a dementia-free person will develop dementia later in life, DFLE' = Dementia-free life expectancy for a dementia-free person of given age.
By construction, DFLE and DLE have the same standard error, and LE has zero variance.

Quantities calculated using fitted values of dementia incidence and relative risk of death (with dementia vs. without) shown in Appendix Table 11.

Subjects who survived ADAMS without a diagnosis of dementia were censored at the ADAMS study period, rather than at their last assessment.
Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For relative risk of death, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1920 birth cohort. Mortality by dementia status data from Health and Retirement Study (HRS), United States, 2001-2011.

**Table A.2.5 e: Life cycle quantities for dementia, non-stationary approach, censoring subjects without dementia at end of ADAMS study period, 1940 cohort life table**

A) Males

| Age | LE | DFLE | DLE | (SE) | Lifetime Risk | (SE) | DFLE' | (SE) |
|---|---|---|---|---|---|---|---|---|
| 70 | 13.64 | 12.88 | 0.76 | (0.222) | 0.289 | (0.036) | 13.28 | (0.193) |
| 75 | 10.65 | 9.95 | 0.70 | (0.223) | 0.282 | (0.036) | 10.47 | (0.139) |
| 80 | 7.96 | 7.35 | 0.62 | (0.207) | 0.270 | (0.041) | 7.91 | (0.097) |
| 85 | 5.70 | 5.18 | 0.52 | (0.181) | 0.253 | (0.050) | 5.71 | (0.061) |
| 90 | 4.05 | 3.59 | 0.46 | (0.156) | 0.239 | (0.062) | 4.05 | (0.047) |
| 95 | 2.95 | 2.50 | 0.45 | (0.147) | 0.235 | (0.076) | 2.90 | (0.049) |
| 100 | 2.30 | 1.77 | 0.52 | (0.167) | 0.247 | (0.096) | 2.20 | (0.061) |

B) Females

| Age | LE | DFLE | DLE | (SE) | Lifetime Risk | (SE) | DFLE' | (SE) |
|---|---|---|---|---|---|---|---|---|
| 70 | 15.99 | 14.41 | 1.58 | (0.288) | 0.349 | (0.040) | 14.85 | (0.269) |
| 75 | 12.62 | 11.11 | 1.51 | (0.294) | 0.340 | (0.043) | 11.80 | (0.194) |
| 80 | 9.57 | 8.22 | 1.35 | (0.285) | 0.325 | (0.049) | 9.01 | (0.145) |
| 85 | 6.93 | 5.80 | 1.13 | (0.261) | 0.304 | (0.059) | 6.57 | (0.098) |
| 90 | 4.89 | 3.97 | 0.91 | (0.229) | 0.283 | (0.070) | 4.65 | (0.063) |
| 95 | 3.49 | 2.71 | 0.77 | (0.202) | 0.271 | (0.084) | 3.29 | (0.051) |
| 100 | 2.63 | 1.88 | 0.76 | (0.202) | 0.276 | (0.103) | 2.44 | (0.058) |

LE = total life expectancy for a randomly chosen person in the population of given age. DFLE = Dementia-free life expectancy for a randomly chosen person in the population of given age, DLE = life expectancy with dementia for a randomly chosen person in the population of given age, Lifetime Risk = probability that a dementia-free person will develop dementia later in life, DFLE' = Dementia-free life expectancy for a dementia-free person of given age.
By construction, DFLE and DLE have the same standard error, and LE has zero variance.
Quantities calculated using fitted values of dementia incidence and relative risk of death (with dementia vs. without) shown in Appendix Table 11.
Subjects who survived ADAMS without a diagnosis of dementia were censored at the end of the ADAMS study period, rather than at their last assessment.
Data: Aging, Demographics, and Memory Study (ADAMS), United States, 2001-2009. For relative risk of death, n=856. For incidence, n=456. Mortality data from United States Social Security Administration life tables for 1940 birth cohort. Mortality by dementia status data from Health and Retirement Study (HRS), United States, 2001-2011.

## A.3. Appendices to Chapter 3 (Diabetes and mobility limitations)

### A.3.1. Additional sample characteristics

**Table A.3.1 a: Weighted means or proportions of characteristics of individuals who had diabetes at baseline vs. individuals included in propensity score matching procedure**

| | Diabetic at baseline (excluded) | All included individuals | Sig.† |
|---|---|---|---|
| Mother's education (years) | 8.886 | 9.837 | *** |
| Father's education (years) | 8.367 | 9.470 | *** |
| Childhood health (1=excellent) | 1.904 | 1.758 | ** |
| Age (last birthday) | 55.029 | 54.483 | *** |
| Female | 47.1% | 50.6% | |
| Black | 20.1% | 9.7% | *** |
| Hispanic | 9.4% | 6.6% | *** |
| Foreign born | 9.1% | 8.7% | |
| Own education (years) | 11.9 | 12.8 | *** |
| Less than high school | 29.3% | 18.0% | *** |
| High school graduate | 35.1% | 37.1% | * |
| Some college | 19.4% | 22.1% | ** |
| College graduate | 16.2% | 22.7% | *** |
| Northeast | 18.6% | 19.4% | |
| Midwest | 26.1% | 24.8% | |
| South | 41.7% | 37.2% | *** |
| West | 13.6% | 18.6% | *** |
| Married | 75.7% | 80.5% | * |
| Divorced | 24.1% | 19.5% | * |
| Uninsured | 17.4% | 17.3% | |
| Works full time | 44.4% | 62.2% | *** |
| Works part-time | 7.4% | 10.0% | ** |
| Unemployed | 2.6% | 2.0% | |
| Retired | 23.5% | 13.0% | *** |
| Not in labor force | 10.5% | 9.4% | |
| On work disability | 11.5% | 3.5% | *** |
| HH income per head ($) | 18088 | 27119 | *** |
| Wealth ($) | 128368 | 230037 | *** |
| Smokes now | 22.0% | 26.1% | ** |
| Ever smoked | 67.0% | 62.2% | * |

| | | | |
|---|---|---|---|
| BMI ($kg/m^2$) | 30.89 | 27.09 | *** |
| Self-rated health (1=excellent) | 3.572 | 2.392 | *** |
| Has high blood pressure | 60.0% | 28.0% | *** |
| Has had stroke | 7.4% | 1.7% | *** |
| Has arthritis | 46.2% | 30.3% | *** |
| # Mobility limitations (max=5) | 1.697 | 0.784 | *** |
| Number of subjects | 1094 | 9646 | |

Source: Author's calculations from Health and Retirement Study (United States, 1992-2010).

† Significance tests: Pearson Chi-Squared test for differences between groups (proportions); Wilcoxon rank-sum test for differences between groups (sample means): *p<0.05, **p<0.01, ***p<0.001.

**Table A.3.1 b: Diabetes incidence and survival in Original HRS and War Babies cohorts**

**Original HRS cohort (born 1931-1941)**

| Year | Wave | Respondents | Diabetic | At risk | New cases | New deaths | Censored | No response |
|---|---|---|---|---|---|---|---|---|
| 1992 | 1 | 8006 | 0 | 8006 | 0 | 0 | 0 | 0 |
| 1994 | 2 | 7639 | 0 | 7639 | 135 | 0 | 0 | 367 |
| 1996 | 3 | 7575 | 129 | 7446 | 168 | 0 | 0 | 431 |
| 1998 | 4 | 7235 | 269 | 6966 | 154 | 134 | 135 | 771 |
| 2000 | 5 | 6865 | 379 | 6486 | 190 | 187 | 240 | 1141 |
| 2002 | 6 | 6590 | 521 | 6069 | 227 | 249 | 314 | 1416 |
| 2004 | 7 | 6317 | 687 | 5630 | 225 | 189 | 254 | 1689 |
| 2006 | 8 | 5994 | 833 | 5161 | 214 | 248 | 407 | 2012 |
| 2008 | 9 | 5712 | 961 | 4751 | 180 | 262 | 299 | 2294 |
| 2010 | 10 | 5225 | 1006 | 4219 | 206 | 367 | 402 | 2781 |

**War Babies cohort (born 1942-1947)**

| Year | Wave | Respondents | Diabetic | At risk | New cases | New deaths | Censored | No response |
|---|---|---|---|---|---|---|---|---|
| 1998 | 4 | 1642 | 0 | 1642 | 0 | 0 | 0 | 0 |
| 2000 | 5 | 1583 | 0 | 1583 | 31 | 0 | 0 | 59 |
| 2002 | 6 | 1593 | 30 | 1563 | 60 | 0 | 0 | 49 |
| 2004 | 7 | 1534 | 83 | 1451 | 39 | 12 | 12 | 108 |
| 2006 | 8 | 1482 | 116 | 1366 | 54 | 26 | 27 | 160 |
| 2008 | 9 | 1441 | 159 | 1282 | 51 | 25 | 27 | 201 |
| 2010 | 10 | 1369 | 194 | 1175 | 66 | 27 | 29 | 273 |

Diabetic = Number of subjects who had already reported having diabetes in a previous wave.
At risk = Number of subjects at risk of incident diabetes this wave.
New cases = Number of diabetes cases first observed this wave.
New deaths = Number of new deaths reported this wave.
Censored = Number of subjects censored this wave = Number of new deaths + Number newly dropped from survey.
No response = Number of subjects who did not respond to survey this wave.

Source: Author's calculations from Health and Retirement Study (United States, 1992-2010).

Number of incident cases in row for wave *w* equals the number of diabetes cases first observed at wave *w*, i.e. the number of incident cases that occurred between wave *w-1* and wave *w*. Number already diabetic in wave *w* is approximately the sum of all incident cases prior to wave *w*. Without sample attrition, nonresponse, and mortality, it would be exactly equal. Data were collected every two years; calculating an incidence rate requires additional knowledge or assumptions about when in the interval people died or were lost to follow-up. # censored = # who died since previous wave + # dropped from sample since previous wave. In general, individuals who did not respond in a given wave remained in the sample, and the HRS data collection team attempted to reach them in subsequent waves. Individuals who died before their cohort was observed three times were excluded from the sample because they did not provide enough data to identify an effect of incident diabetes on subsequent disability.

### A.3.2. Propensity score model fit and matching results

**Table A.3.2 a: Propensity score model fit**

Dependent variable = Diagnosed with diabetes

| Independent variable | Odds ratio | 95% Conf. Int. |
|---|---|---|
| Mother's education (years) | 0.988 | [0.968,1.009] |
| Father's education (years) | 0.994 | [0.974,1.015] |
| Childhood health [b] | 0.971 | [0.916,1.031] |
| Age (last birthday) | 1.010 | [0.996,1.024] |
| Female | 0.765*** | [0.677,0.864] |
| Black | 1.231 | [0.996,1.523] |
| Hispanic | 1.452** | [1.167,1.806] |
| Foreign born | 0.995 | [0.753,1.314] |
| Own education (years) | 0.987 | [0.968,1.006] |
| Midwest | 0.924 | [0.765,1.116] |
| Northeast | 0.928 | [0.816,1.055] |
| West | 0.999 | [0.868,1.149] |
| Divorced | 0.915 | [0.769,1.089] |
| Uninsured | 1.197 | [0.972,1.473] |
| Works full-time | 1.130 | [0.976,1.308] |
| Wealth (Z-score) [a] | 0.949 | [0.847,1.064] |
| Household income per head (Z-score) [a] | 0.511* | [0.267,0.977] |
| Current smoker | 1.034 | [0.859,1.245] |
| Former smoker | 0.977 | [0.827,1.155] |
| BMI ($kg/m^2$) | 1.101*** | [1.086,1.116] |
| Self-rated health [b] | 1.177*** | [1.101,1.259] |
| High blood pressure | 1.641*** | [1.477,1.823] |
| Heart problems | 1.254** | [1.077,1.460] |
| Stroke | 1.071 | [0.833,1.378] |
| Arthritis | 0.923 | [0.778,1.094] |
| Number mobility limitations (max=5) | 0.980 | [0.927,1.036] |
| 1994-1996 interval | 1.327* | [1.047,1.682] |
| 1996-1998 interval | 1.270 | [0.936,1.722] |
| 1998-2000 interval | 1.330* | [1.066,1.659] |
| 2000-2002 interval | 2.060*** | [1.675,2.533] |
| 2002-2004 interval | 1.776*** | [1.381,2.284] |
| 2004-2006 interval | 2.088*** | [1.555,2.804] |
| 2006-2008 interval | 1.953*** | [1.427,2.673] |
| 2008-2010 interval | 2.429*** | [1.735,3.401] |
| *N (person-waves)* | 64776 | |

(Continued on next page)

Notes: Author's calculations from Health and Retirement Study (United States, 1992-2010).
[*] $p < 0.05$, [**] $p < 0.01$, [***] $p < 0.001$.
[a] Household wealth and income per head were measured in standard-deviation units, so their odds ratios show the proportionate increased odds of developing diabetes associated with a one-standard-deviation increase in wealth or income per head, respectively. Sample mean wealth across all person-observations was $333,774 with standard deviation $1,285,870. Income per head sample mean was $30,359 with standard deviation $227,406.
[b] Childhood health (self-rated) and current self-rated health were measured on a scale from 1 to 5, with 1 indicating the best health.

**Table A.3.2 b: Balance of covariates before and after risk-set matching**

| When measured | Variable | Std diff before matching | Std diff after matching | Exposed mean before | Control mean before | P before | Exposed mean after | Control mean after | P after |
|---|---|---|---|---|---|---|---|---|---|
| Time-invariant | Mother's education (yrs) | -0.171 | -0.008 | 8.695 | 9.6 | <0.001 | 8.8 | 8.8 | 0.980 |
| | Father's education (yrs) | -0.147 | 0.010 | 8.378 | 9.2 | <0.001 | 8.5 | 8.4 | 0.607 |
| | Childhood health | 0.048 | 0.047 | 1.868 | 1.800 | 0.010 | 1.865 | 1.799 | 0.240 |
| | female | -0.046 | -0.019 | 0.500 | 0.532 | 0.010 | 0.501 | 0.514 | 0.458 |
| | Own education (yrs) | -0.181 | -0.010 | 11.7 | 12.5 | <0.001 | 11.8 | 11.9 | 0.914 |
| | Black | 0.138 | 0.053 | 0.215 | 0.140 | <0.001 | 0.209 | 0.180 | 0.040 |
| | Hispanic | 0.126 | 0.081 | 0.128 | 0.074 | <0.001 | 0.120 | 0.086 | 0.001 |
| | Foreign born | 0.062 | 0.090 | 0.121 | 0.094 | <0.001 | 0.115 | 0.076 | <0.001 |

(Table continues on next three pages.)

| When measured | Variable | Std diff before matching | Std diff after matching | Exposed mean before | Control mean before | P before | Exposed mean after | Control mean after | P after |
|---|---|---|---|---|---|---|---|---|---|
| Baseline (study entry) | Age | -0.014 | -0.020 | 54.95 | 55.01 | 0.541 | 54.97 | 55.06 | 0.280 |
| | South | 0.066 | -0.067 | 0.451 | 0.405 | <0.001 | 0.440 | 0.487 | 0.008 |
| | West | -0.028 | 0.070 | 0.153 | 0.167 | 0.120 | 0.156 | 0.120 | 0.003 |
| | Midwest | -0.020 | -0.010 | 0.235 | 0.247 | 0.252 | 0.232 | 0.238 | 0.677 |
| | Northeast | -0.036 | 0.032 | 0.162 | 0.181 | 0.046 | 0.172 | 0.155 | 0.198 |
| | Married | -0.016 | -0.118 | 0.816 | 0.824 | 0.354 | 0.808 | 0.873 | <0.001 |
| | Divorced | 0.016 | 0.118 | 0.185 | 0.176 | 0.254 | 0.192 | 0.127 | <0.001 |
| | Household income ($) | -0.151 | -0.037 | 18,480 | 24,441 | <0.001 | 18,767 | 20,239 | 0.025 |
| | Wealth ($) | -0.137 | -0.025 | 143,552 | 224,561 | <0.001 | 145,068 | 159,926 | 0.002 |
| | Works full time | -0.045 | 0.005 | 0.570 | 0.601 | 0.011 | 0.571 | 0.567 | 0.830 |
| | Works part time | -0.017 | -0.046 | 0.097 | 0.104 | 0.336 | 0.094 | 0.114 | 0.073 |
| | Unemployed | 0.008 | 0.006 | 0.025 | 0.023 | 0.651 | 0.026 | 0.024 | 0.821 |
| | Retired | 0.020 | 0.000 | 0.149 | 0.139 | 0.252 | 0.151 | 0.151 | 1.000 |
| | On disability | 0.050 | 0.014 | 0.047 | 0.033 | 0.003 | 0.046 | 0.042 | 0.607 |
| | Not in labor force | 0.031 | 0.024 | 0.114 | 0.100 | 0.080 | 0.112 | 0.101 | 0.330 |
| | BMI ($kg/m^2$) | 0.497 | 0.035 | 29.86 | 26.31 | <0.001 | 29.62 | 29.37 | 0.082 |
| | Smokes now | 0.006 | 0.088 | 0.271 | 0.267 | 0.735 | 0.272 | 0.217 | <0.001 |
| | Former smoker | -0.004 | -0.037 | 0.354 | 0.357 | 0.814 | 0.358 | 0.383 | 0.143 |
| | Uninsured | 0.061 | 0.047 | 0.223 | 0.188 | 0.001 | 0.215 | 0.188 | 0.058 |
| | Self-rated health | 0.216 | 0.035 | 2.730 | 2.379 | <0.001 | 2.699 | 2.642 | 0.246 |
| | High blood pressure | 0.254 | -0.006 | 0.437 | 0.267 | <0.001 | 0.423 | 0.426 | 0.830 |
| | Heart disease | 0.060 | 0.014 | 0.101 | 0.077 | <0.001 | 0.100 | 0.094 | 0.591 |
| | Stroke | 0.014 | 0.061 | 0.021 | 0.018 | 0.420 | 0.024 | 0.012 | 0.013 |

| When measured | Variable | Std diff before matching | Std diff after matching | Exposed mean before | Control mean before | P before | Exposed mean after | Control mean after | P after |
|---|---|---|---|---|---|---|---|---|---|
| Baseline (Study entry) | Arthritis | 0.079 | -0.017 | 0.355 | 0.303 | <0.001 | 0.352 | 0.363 | 0.507 |
| | # Mobility limitations | 0.202 | 0.063 | 1.203 | 0.822 | <0.001 | 1.200 | 1.082 | 0.017 |
| Just before diabetes onset | Age | 0.033 | -0.014 | 61.98 | 61.74 | 0.637 | 62.34 | 62.44 | 0.687 |
| | South | 0.128 | -0.075 | 0.452 | 0.363 | <0.001 | 0.454 | 0.506 | 0.003 |
| | West | -0.004 | 0.079 | 0.148 | 0.150 | 0.826 | 0.157 | 0.117 | 0.001 |
| | Midwest | 0.008 | -0.010 | 0.220 | 0.215 | 0.644 | 0.226 | 0.232 | 0.705 |
| | Northeast | 0.005 | 0.032 | 0.152 | 0.149 | 0.765 | 0.162 | 0.146 | 0.203 |
| | Married | 0.115 | -0.124 | 0.799 | 0.730 | <0.001 | 0.814 | 0.888 | <0.001 |
| | Divorced | 0.045 | 0.140 | 0.171 | 0.148 | 0.010 | 0.184 | 0.112 | <0.001 |
| | Household income ($) | -0.121 | 0.007 | 21661 | 33022 | <0.001 | 22285 | 21660 | 0.108 |
| | Wealth ($) | -0.113 | 0.001 | 203585 | 339099 | <0.001 | 216712 | 215473 | 0.008 |
| | Works full time | 0.006 | 0.003 | 0.319 | 0.315 | 0.722 | 0.320 | 0.318 | 0.909 |
| | Works part time | -0.018 | -0.018 | 0.058 | 0.064 | 0.329 | 0.058 | 0.064 | 0.461 |
| | Unemployed | 0.020 | 0.019 | 0.011 | 0.008 | 0.247 | 0.011 | 0.009 | 0.477 |
| | Retired | 0.091 | -0.014 | 0.443 | 0.379 | <0.001 | 0.466 | 0.476 | 0.571 |
| | On disability | 0.062 | 0.040 | 0.051 | 0.033 | <0.001 | 0.052 | 0.041 | 0.132 |
| | Not in labor force | 0.030 | 0.002 | 0.092 | 0.080 | 0.085 | 0.092 | 0.092 | 0.951 |
| | BMI ($kg/m^2$) | 0.106 | 0.014 | 30.941 | 30.080 | <0.001 | 30.640 | 30.526 | 0.410 |
| | Smokes now | -0.021 | 0.051 | 0.181 | 0.193 | 0.251 | 0.180 | 0.152 | 0.033 |
| | Former smoker | 0.086 | 0.003 | 0.435 | 0.375 | <0.001 | 0.454 | 0.452 | 0.915 |
| | Uninsured | 0.070 | 0.067 | 0.136 | 0.104 | <0.001 | 0.130 | 0.100 | 0.007 |
| | Self-rated health | 0.302 | 0.034 | 3.031 | 2.568 | <0.001 | 3.000 | 2.948 | 0.201 |
| | High blood pressure | 0.318 | -0.017 | 0.614 | 0.395 | <0.001 | 0.610 | 0.622 | 0.490 |

| When measured | Variable | Std diff before matching | Std diff after matching | Exposed mean before | Control mean before | P before | Exposed mean after | Control mean after | P after |
|---|---|---|---|---|---|---|---|---|---|
| Just before diabetes onset | Heart disease | 0.148 | 0.070 | 0.215 | 0.136 | <0.001 | 0.220 | 0.182 | 0.008 |
| | Stroke | 0.042 | 0.098 | 0.053 | 0.040 | 0.016 | 0.059 | 0.030 | <0.001 |
| | Arthritis | 0.104 | -0.021 | 0.554 | 0.481 | <0.001 | 0.553 | 0.568 | 0.393 |
| | # Mobility limitations | 0.239 | 0.060 | 1.156 | 0.710 | <0.001 | 1.138 | 1.026 | 0.057 |

Std diff = Standardized difference = ((exposed group mean) – (control mean))/(Pooled standard deviation). BMI = Body Mass Index.
Childhood health and self rated health: 1=excellent, 5 is the worst possible health.
Household income is household income per person living in the household.
P-values are from Pearson chi-squared tests for binary variables (e.g. “smokes now”), and Wilcoxon rank-sum tests for numeric variables (e.g. household income per head). Small P-values indicate a statistically significant difference in means between the exposed and control group.

### A.3.3. Results using alternative specifications

To test the sensitivity of the results to the restriction that only diagnosed diabetes is observed, matched pairs in which the control unit was diagnosed with diabetes in some later wave, after being matched, were dropped. Individuals subsequently diagnosed with diabetes are probably more likely than those who are never diagnosed to have had undiagnosed diabetes when matched as controls. Because those with undiagnosed diabetes might experience some of the disabling effects of the disease, removing pairs with control units who later got diabetes could increase the within-pair differences in disability accumulation, increasing the magnitude of the estimated association between incident diabetes and mobility limitation. In fact, as shown in **Appendix Table 5**, the results are almost identical when these pairs are excluded, which suggests that ignoring undiagnosed diabetes does not produce a major bias in either direction.

**Table A.3.3 a: Signed-rank test results, excluding pairs with control individual who later developed diabetes**

(Number of pairs = 1323)

| | |
|---|---|
| Hodges-Lehman estimate | 0.25 |
| 95% CI for H-L estimate | [0.000009, 0.25001] |
| p-value for $H_0$: No difference between exposed and matched-controls | 0.0001 |

Notes: Author's calculations from Health and Retirement Study (United States, 1992-2010).

Hodges-Lehman estimate is the median of pairwise Walsh averages; it can be interpreted roughly as the median within-pair difference between an exposed and control unit.

**Table A.3.3 b: Comparing mobility change among incident cases of diabetes to mobility change among matched controls, defining number of mobility limitations at diagnosis as number of mobility limitations at first report of diabetes diagnosis**

(Number of pairs = 1559)

| | |
|---|---|
| p-value for $H_0$: No difference between exposed and matched-controls | 0.000001557 |
| Hodges-Lehman estimate: median of pairwise Walsh averages (approximately, the median within-pair difference) | 0.2499 |
| 95% CI for H-L estimate | [0.2499, 0.2500] |

Source: Author's calculations from Health and Retirement Study (United States, 1992-2010).

### A.3.4. Sensitivity to unmeasured confounding

The Wilcoxon signed-rank test and its associated Hodges-Lehman estimate represent a test and estimate of the effect of incident diabetes on subsequent mobility limitation, under the assumption of no unmeasured confounding (Rosenbaum 2002). By observing incident diabetes and subsequent disabilities and ensuring balance between exposed (incident-diabetes) and control groups prior to diabetes onset, this paper reduces the potential confounding found in other studies. However, there could still be unmeasured forces that affect both diabetes onset above age 50 and subsequent mobility limitation. This appendix estimates the magnitude of unmeasured confounding (also called hidden bias) that would be needed to explain any observed association between incident diabetes and subsequent mobility limitation.

The claim underlying matched-sets inference is that within matched sets, i.e. conditional on all the covariates used in the matching process, the probability of exposure for any unit in the set is the same. When dealing with matched pairs, that probability is ½. The signed-rank test then assesses whether the difference in outcomes between the exposed and control units could plausibly have resulted from random sampling error, given that each unit in the pair had a probability of exposure of ½. The sensitivity analysis addresses how inferences would be altered if unmeasured confounders, which have arbitrarily strong associations with the outcome, increased the probability of exposure for one unit in each pair (Gastwirth, Krieger, & Rosenbaum 1998; Rosenbaum 1987; Silber et al. 2009). The procedure estimates the magnitude of bias needed to explain the study's observed findings by simulating Wilcoxon signed-rank tests for

within-pair probabilities of exposure other than ½. The procedure was designed for use in studies based on propensity-score matching and detailed by Rosenbaum (2002).

*Results*

**Table A.3.4 a** shows the results of a simulation of different magnitudes of hidden bias, that is, unmeasured variables that are arbitrarily highly associated with mobility limitation and raise the probability of developing diabetes for one unit in each matched pair. A research finding unlikely to have resulted from unmeasured confounding is one in which the maximum p-value remains below a specified threshold (conventionally 0.05) even when the magnitude of hidden bias, represented by Gamma (Γ), is fairly large. This table shows that if some unmeasured confounder that was highly positively associated with mobility limitations increased the probability of developing diabetes by 15%, the significant Hodges-Lehman result shown in **Table 3-3** could be completely explained by this unmeasured confounder.

**Table A.3.4 a: Sensitivity of signed-rank test results to unmeasured confounding**

| Gamma (Γ) | Minimum p-value | Maximum p-value |
|---|---|---|
| 1 | 0.000074 | 0.000074 |
| 1.1 | 0.000000034 | 0.014 |
| 1.15 | 0.00000000040 | 0.072 |
| 1.2 | 0.0000000000033 | 0.23 |

Gamma (Γ) is a measure of the magnitude of unmeasured confounding (hidden bias), where Γ=1 means no unmeasured confounding. For all values of Γ, the association between the unmeasured confounder and the outcome is arbitrarily strong and positive. Minimum and maximum p-values are for Wilcoxon signed-rank test of the null hypothesis of no difference in change in mobility limitations between exposed and matched control units. A range of p-values is produced because the results are based on simulating alternative probabilities of exposure, based on the magnitude of the hidden bias.

### A.3.5. Results when considering only pairs in which both units were obese at baseline

**Table A.3.5 a: Signed-rank test results, pairs in which both units were obese at baseline**

(Number of pairs = 466)

| | |
|---|---|
| p-value for $H_0$: No difference between exposed and matched-controls | 0.02983 |
| Hodges-Lehman estimate: median of pairwise Walsh averages (approximately, the median within-pair difference) | 0.25 |
| 95% CI for H-L estimate | [0.00003196, 0.49996307] |

Source: Author's calculations from Health and Retirement Study (United States, 1992-2010).

**Table A.3.5 b: Sensitivity of results to hidden bias, pairs in which both units were obese at baseline**

| Gamma (Γ) | Minimum p-value | Maximum p-value |
|---|---|---|
| 1 | 0.01577 | 0.01577 |
| 1.05 | 0.004686 | 0.04425 |
| 1.06 | 0.003623 | 0.05298 |
| 1.1 | 0.001236 | 0.1006 |

Source: Author's calculations from Health and Retirement Study (United States, 1992-2010).

Gamma is a measure of the magnitude of hidden bias, where Γ=1 means no hidden bias. Minimum and maximum p-values are for Wilcoxon signed-rank test of the null hypothesis of no difference in disability change between exposed and matched control units. A range of p-values is produced because the results are based on simulating alternative probabilities of exposure, based on the magnitude of the hidden bias.

**Table A.3.5 c: Differences in average number of mobility limitations, both units in pair obese at baseline**

(Number of pairs = 466)

| Time | Exposed mean | Exposed SD | Control mean | Control SD |
|---|---|---|---|---|
| Baseline | 1.562 | (1.531529) | 1.519 | (1.49862) |
| Last observation pre-diagnosis | 1.479 | (1.542436) | 1.386 | (1.438691) |
| Censoring time | 2.152 | (1.722827) | 1.880 | (1.667701) |
| Average change | 0.632 | | 0.427 | |

Average change = (# mobility limitations at censoring time) – [(# at baseline)+(# at time t)]/2. See Equation 3 in text.

Time *t* = last observation before diabetes onset in exposed individual

### A.3.6. Association between duration of diabetes exposure and mobility limitations

To assess the association between duration of time since initial exposure and mobility limitations at censoring time, against the alternative of not being exposed at the initial exposure time (but possibly being exposed at subsequent times), I include length of follow-up as a covariate in my Poisson model along with an interaction between exposed status (exposed or control) and length of follow-up. The rate ratio for the interaction term represents how many more mobility limitations are expected for the exposed person for an additional unit of follow-up. Median length of follow-up following diabetes diagnosis was three waves (six years), with a maximum of nine waves (18 years). Results are shown with and without fixed effects for the total number of mobility limitations in each matched pair; the fixed effects model is the preferred model (Lachin 2011).

The results provide no evidence for the importance of duration. The rate ratio for the interaction term was approximately 1 and not statistically significant in either model. The predominance of individuals with short follow-up periods (i.e. few years after diabetes diagnosis) and the exclusion of individuals who had diabetes at study entry are likely explanations for this null finding.

**Table A.3.6 a: Poisson regression when incorporating length of follow-up**

Dependent variable = Number of mobility limitations at censoring time

| Independent variable | (1) No fixed effects | (2) Fixed effects for total # mobility limitations at censoring time in matched pair |
|---|---|---|
| exposed | 1.238*** | 1.238*** |
| | [1.109,1.381] | [1.109,1.380] |
| follow-up length | 1.023* | 1.003 |
| | [1.004,1.042] | [0.985,1.022] |
| interaction | 0.994 | 0.994 |
| | [0.969,1.019] | [0.969,1.019] |
| N | 3172 | 2534 |

Notes: Author's calculations from Health and Retirement Study (United States, 1992-2010).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Model (2) dropped observations in matched pairs with a total number of mobility limitations of zero.

Point estimate is the incidence rate ratio, which is the exponentiated coefficient. 95% Confidence Intervals shown.

Each unit of follow-up length is one wave of data collection, corresponding to two years.

*Interaction* is the product of *exposed* and *follow-up length*.

## References


Abatih, E., Van Oyen, H., Bossuyt, N., & Bruckers, L. (2008). Variance estimation methods for health expectancy by relative socio-economic status. *European Journal of Epidemiology*, *23*, 243–249. doi:10.1007/s10654-008-9227-5

Abdullah, A., Stoelwinder, J., Shortreed, S., Wolfe, R., Stevenson, C., Walls, H., … Peeters, A. (2011). The duration of obesity and the risk of type 2 diabetes. *Public Health Nutr*, *14*(1), 119–126. doi:10.1017/S1368980010001813

Agüero-Torres, H., Fratiglioni, L., & Guo, Z. (1999). Mortality from dementia in advanced age: a 5-year follow-up study of incident dementia cases. *Journal of Clinical Epidemiology*, *52*(8), 737–743. Retrieved from http://www.sciencedirect.com/science/article/pii/S0895435699000670

Ali, M. K., Bullard, K. M., Saaddine, J. B., Cowie, C. C., Imperatore, G., & Gregg, E. W. (2013). Achievement of goals in U.S. diabetes care, 1999-2010. *N Engl J Med*, *368*(17), 1613–1624. doi:10.1056/NEJMsa1213829

Allison, P. D. (1984). *Event History Analysis*. Beverly Hills, CA: SAGE Publications.

Allison, P. D. (2002). *Missing Data*. Thousand Oaks, CA: SAGE Publications.

Alzheimer's Association. (2014). 2014 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *10*(2), e47–e92. doi:10.1016/j.jalz.2014.02.001

American Diabetes Association. (2012). Diagnosis and classification of diabetes mellitus. *Diabetes Care*, *35 Suppl 1*, S64–71. doi:10.2337/dc12-s064

American Diabetes Association. (2013). Economic costs of diabetes in the U.S. in 2012. *Diabetes Care*, *36*(4), 1033–46. doi:10.2337/dc12-2625

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision*. Arlington, VA.

Asgharian, M., Wolfson, D. B., & Zhang, X. (2006). Checking stationarity of the incidence rate using prevalent cohort survival data. *Statistics in Medicine*, *25*(10), 1751–1767. doi:10.1002/sim.2326

Bardenheier, B. H., Gregg, E. W., Zhuo, X., Cheng, Y. J., & Geiss, L. S. (2014). Association of Functional Decline With Subsequent Diabetes Incidence in U.S. Adults Aged 51 Years and Older: The Health and Retirement Study 1998-2010. *Diabetes Care*, 1–7. doi:10.2337/dc13-2216

Bell, F. C., & Miller, M. L. (2005). Life Tables for the United States Social Security Area 1900-2100. *Actuarial Study No. 120*. Retrieved from http://www.ssa.gov/oact/NOTES/as120/LifeTables_Tbl_7.html

Berger, B., Stenström, G., & Sundkvist, G. (1999). Incidence, prevalence, and mortality of diabetes in a large population. A report from the Skaraborg Diabetes Registry.

*Diabetes Care*, *22*(5), 773–778.

Bertoni, A. G., & Goff, D. C. J. (2011). Diabetes and Cardiovascular Disease. In K. M. V. Narayan, D. E. Williams, E. W. Gregg, & C. C. Cowie (Eds.), *Diabetes Public Health* (pp. 173–194). New York, NY: Oxford University Press.

Bianchi, L., Zuliani, G., & Volpato, S. (2013). Physical disability in the elderly with diabetes: epidemiology and mechanisms. *Current Diabetes Reports*, *13*(6), 824–30. doi:10.1007/s11892-013-0424-6

Bonora, E., & Tuomilehto, J. (2011). The pros and cons of diagnosing diabetes with A1C. *Diabetes Care*, *34 Suppl 2*, S184–90. doi:10.2337/dc11-s216

Boyle, J. P., Thompson, T. J., Gregg, E. W., Barker, L. E., & Williamson, D. F. (2010). Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. *Population Health Metrics*, *8*, 29. doi:10.1186/1478-7954-8-29

Brookmeyer, R., Evans, D. A., Hebert, L., Langa, K. M., Heeringa, S. G., Plassman, B. L., & Kukull, W. A. (2011). National estimates of the prevalence of Alzheimer's disease in the United States. *Alzheimer's & Dementia*, *7*(1), 61–73. doi:10.1016/j.jalz.2010.11.007

Brookmeyer, R., & Gray, S. (2000). Methods for projecting the incidence and prevalence of chronic diseases in aging populations: application to Alzheimer's disease. *Statistics in Medicine*, *19*(11-12), 1481–93. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10844713

Brown, H. S., Pagán, J. A., & Bastida, E. (2005). The impact of diabetes on employment: genetic IVs in a bivariate probit. *Health Economics*, *14*(5), 537–44. doi:10.1002/hec.942

Bruce, D. G., Davis, W. A., & Davis, T. M. E. (2005). Longitudinal Predictors of Reduced Mobility and Physical Disability in Patients With Type 2 Diabetes: The Fremantle Diabetes Study. *Diabetes Care*, *28*(10), 2441–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16186277

Bullard, K. M., Saydah, S. H., Imperatore, G., Cowie, C. C., Gregg, E. W., Geiss, L. S., … Caspersen, C. J. (2013). Secular Changes in U.S. Prediabetes Prevalence Defined by Hemoglobin A1c and Fasting Plasma Glucose: National Health and Nutrition Examination Surveys, 1999-2010. *Diabetes Care*, *36*(8), 2286–2293. doi:10.2337/dc12-2563

Carone, M., Asgharian, M., & Jewell, N. P. (2014). Estimating the Lifetime Risk of Dementia in the Canadian Elderly Population Using Cross-Sectional Cohort Survival Data. *Journal of the American Statistical Association*, *109*(505), 24–35. doi:10.1080/01621459.2013.859076

Caruso, L. B., Silliman, R. A., Demissie, S., Greenfield, S., & Wagner, E. H. (2000).

What can we do to improve physical function in older persons with type 2 diabetes? *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *55*(7), M372–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10898253

Centers for Disease Control and Prevention. (2012). Crude and Age-Adjusted Incidence of Diagnosed Diabetes per 1,000 Population Aged 18-79 Years, United States, 1980-2007. (U.S. Department of Health and Human Services, Ed.). Atlanta, GA. Retrieved from http://www.cdc.gov/diabetes/statistics/incidence/fig2.htm

Centers for Disease Control and Prevention. (2014). Compressed Mortality Files, 1979-1998 and 1999-2013. Hyattsville, MD: National Center for Health Statistics: CDC WONDER. Retrieved from http://wonder.cdc.gov/cmf-icd10.html

Chêne, G., Beiser, A., Au, R., Preis, S. R., Wolf, P. a, Dufouil, C., & Seshadri, S. (2015). Gender and incidence of dementia in the Framingham Heart Study from mid-adult life. *Alzheimer's & Dementia*, *11*(3), 310–320. doi:10.1016/j.jalz.2013.10.005

Cheng, C., Lin, C.-H., Tsai, Y.-W., Tsai, C.-J., Chou, P.-H., & Lan, T.-H. (2014). Type 2 Diabetes and Antidiabetic Medications in Relation to Dementia Diagnosis. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, (12), 1–7. doi:10.1093/gerona/glu073

Cheng, Y. J., Imperatore, G., Geiss, L. S., Wang, J., Saydah, S. H., Cowie, C. C., & Gregg, E. W. (2013). Secular changes in the age-specific prevalence of diabetes among U.S. adults: 1988-2010. *Diabetes Care*, *36*(9), 2690–6. doi:10.2337/dc12-2074

Chiu, C.-J., & Wray, L. A. (2011). Physical disability trajectories in older Americans with and without diabetes: the role of age, gender, race or ethnicity, and education. *The Gerontologist*, *51*(1), 51–63. doi:10.1093/geront/gnq069

Cigolle, C. T., Lee, P. G., Langa, K. M., Lee, Y.-Y., Tian, Z., & Blaum, C. S. (2011). Geriatric conditions develop in middle-aged adults with diabetes. *Journal of General Internal Medicine*, *26*(3), 272–9. doi:10.1007/s11606-010-1510-y

Cowie, C. C., Rust, K. F., Byrd-Holt, D. D., Gregg, E. W., Ford, E. S., Geiss, L. S., … Fradkin, J. E. (2010). Prevalence of diabetes and high risk for diabetes using A1C criteria in the U.S. population in 1988-2006. *Diabetes Care*, *33*(3), 562–568. doi:10.2337/dc09-1524

Crimmins, E., & Beltrán-Sánchez, H. (2010). Mortality and Morbidity Trends : Is There Compression of Morbidity ? *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *66B*, 75–86. doi:10.1093/geronb/gbq088.Advance

Crimmins, E., Hayward, M., & Hagedorn, A. (2009). Change in disability-free life expectancy for Americans 70 years old and older. *Demography*, *46*(3), 627–646. Retrieved from http://link.springer.com/article/10.1353/dem.0.0070

Deckers, K., van Boxtel, M. P. J., Schiepers, O. J. G., de Vugt, M., Muñoz Sánchez, J. L.,

Anstey, K. J., … Köhler, S. (2014). Target risk factors for dementia prevention: a systematic review and Delphi consensus study on the evidence from observational studies. *International Journal of Geriatric Psychiatry*, n/a–n/a. doi:10.1002/gps.4245

Dhamoon, M. S., Moon, Y. P., Paik, M. C., Sacco, R. L., & Elkind, M. S. V. (2014). Diabetes predicts long-term disability in an elderly urban cohort: the Northern Manhattan Study. *Annals of Epidemiology*, *24*(5), 362–368.e1. doi:10.1016/j.annepidem.2013.12.013

Eberstein, I. W., Nam, C. B., & Heyman, K. M. (2008). Causes of death and mortality crossovers by race. *Biodemography and Social Biology*, *54*(2), 214–228. doi:10.1080/19485565.2008.9989143

Ewbank, D. (2004). From Alzheimer's disease to a demography of chronic disease: the development of demographic synthesis for fitting multistate models. *Population and Development Review*, 63–86. Retrieved from http://www.jstor.org/stable/3401463

Fishman, E. I. (2015). *Variance Estimation for a Complex Life Table Quantity: Disease-Free Life Expectancy* (No. University of Pennsylvania Population Studies Center WPS 15-2). Philadelphia, PA. Retrieved from http://repository.upenn.edu/psc_working_papers/60/

Flegal, K. M., Carroll, M. D., Kit, B. K., & Ogden, C. L. (2012). Prevalence of obesity and trends in the distribution of body mass index among us adults, 1999-2010. *JAMA: The Journal of the American Medical Association*, *307*(5), 491–497. doi:10.1001/jama.2012.39

Freedman, V. A., Spillman, B. C., Andreski, P. M., Cornman, J. C., Crimmins, E. M., Kramarow, E., … Waidmann, T. A. (2013). Trends in late-life activity limitations in the United States: an update from five national surveys. *Demography*, *50*(2), 661–71. doi:10.1007/s13524-012-0167-z

Friedman, E. M., Shih, R. A., Langa, K. M., & Hurd, M. D. (2015). US Prevalence And Predictors Of Informal Caregiving For Dementia. *Health Affairs*, *34*(10), 1637–1641. doi:10.1377/hlthaff.2015.0510

Garcia-Ptacek, S., Farahmand, B., Kåreholt, I., Religa, D., Cuadrado, M. L., & Eriksdotter, M. (2014). Mortality risk after dementia diagnosis by dementia type and underlying factors: a cohort of 15,209 patients based on the Swedish Dementia Registry. *Journal of Alzheimer's Disease*, *41*(2), 467–77. doi:10.3233/JAD-131856

Gastwirth, J., Krieger, A., & Rosenbaum, P. R. (1998). Dual and simultaneous sensitivity analysis for matched pairs. *Biometrika*, *85*(4), 907–920. Retrieved from http://biomet.oxfordjournals.org/content/85/4/907.short

Geiss, L. S., Pan, L., Cadwell, B., Gregg, E. W., Benjamin, S. M., & Engelgau, M. M. (2006). Changes in incidence of diabetes in US adults, 1997–2003. *Am J Prev Med*, *30*(5), 371–377.

Greenland, S., & Rothman, K. J. (2008). Measures of Occurrence. In K. J. Rothman, S. Greenland, & T. L. Lash (Eds.), *Modern Epidemiology* (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Gregg, E. W., Boyle, J. P., Thompson, T. J., Barker, L. E., Albright, A. L., & Williamson, D. F. (2013). Modeling the impact of prevention policies on future diabetes prevalence in the United States: 2010--2030. *Population health metrics*, *11*(1), 18. doi:10.1186/1478-7954-11-18

Gregg, E. W., Cadwell, B. L., Cheng, Y. J., Cowie, C. C., Williams, D. E., Geiss, L., … Vinicor, F. (2004). Trends in the prevalence and ratio of diagnosed to undiagnosed diabetes according to obesity levels in the US. *Diabetes Care*, *27*(12), 2806–2812.

Gregg, E. W., Cheng, Y. J., Saydah, S., Cowie, C., Garfield, S., Geiss, L., & Barker, L. (2012). Trends in death rates among U.S. adults with and without diabetes between 1997 and 2006: findings from the National Health Interview Survey. *Diabetes Care*, *35*(6), 1252–1257. doi:10.2337/dc11-1162

Gregg, E. W., Geiss, L., Zhang, P., Zhuo, X., Williamson, D. F., & Albright, A. L. (2013). Implications of risk stratification for diabetes prevention: the case of hemoglobin A1c. *Am J Prev Med*, *44*(4 Suppl 4), S375–80. doi:10.1016/j.amepre.2012.12.012

Gregg, E. W., Gu, Q., Cheng, Y. J., Narayan, K. M. V., & Cowie, C. C. (2007). Mortality trends in men and women with diabetes, 1971 to 2000. *Annals of Internal Medicine*, *147*(3), 149–155.

Gregg, E. W., Mangione, C. M., Cauley, J. A., Thompson, T. J., Schwartz, A. V., Ensrud, K. E., & Nevitt, M. C. (2002). Diabetes and Incidence of Functional Disability in Older Women. *Diabetes Care*, *25*(1), 61–67. doi:10.2337/diacare.25.1.61

Gregg, E. W., Zhuo, X., Cheng, Y. J., Albright, A. L., Narayan, K. M. V., & Thompson, T. J. (2014). Trends in lifetime risk and years of life lost due to diabetes in the USA, 1985–2011: a modelling study. *The Lancet Diabetes & Endocrinology*, *8587*(14), 1–8. doi:10.1016/S2213-8587(14)70161-5

Guehne, U., Riedel-Heller, S., & Angermeyer, M. C. (2005). Mortality in Dementia. *Neuroepidemiology*, *25*(3), 153–162. doi:10.1159/000086680

Haas, S. A. (2008). Trajectories of functional health: the "long arm" of childhood health and socioeconomic factors. *Social Science & Medicine*, *66*(4), 849–61. doi:10.1016/j.socscimed.2007.11.004

Hansen, B. B., & Klopfer, S. O. (2006). Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*, *15*, 609–627.

He, W., & Larsen, L. J. (2014). Older Americans With a Disability: 2008 − 2012. *U.S. Census Bureau and National Institute on Aging report*. Retrieved from https://www.census.gov/content/dam/Census/library/publications/2014/acs/acs-

29.pdf

Health and Retirement Study. (2010). *Sampling weights revised for Tracker 2.0 and beyond*. Ann Arbor, MI. Retrieved from http://hrsonline.isr.umich.edu/index.php?p=weightinfo

Health and Retirement Study. (2011). Sample Sizes and Response Rates. Retrieved January 20, 2014, from hrsonline.isr.umich.edu/sitedocs/sampleresponse.pdf

Health and Retirement Study. (2013). Public Use Dataset. Ann Arbor, MI: Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Retrieved January 2, 2014, from http://hrsonline.isr.umich.edu/

Heeringa, S. G., & Connor, J. H. (1995). *Technical Description of the Health and Retirement Survey Sample Design*. Ann Arbor, MI. Retrieved from http://hrsonline.isr.umich.edu/sitedocs/userg/HRSSAMP.pdf

Heeringa, S. G., Hurd, M., Langa, K. M., Ofstedal, M. B., Plassman, B. L., Rodgers, W. L., & Weir, D. R. (2009). *Sample Design, Weighting and Analysis for ADAMS.* Ann Arbor, MI. Retrieved from http://hrsonline.isr.umich.edu/sitedocs/userg/ADAMSSampleWeights_Jun2009.pdf

Helmer, C., Joly, P., Letenneur, L., Commenges, D., & Dartigues, J. F. (2001). Mortality with dementia: Results from a French prospective community-based cohort. *American Journal of Epidemiology*, *154*(7), 642–648. doi:10.1093/aje/154.7.642

Henning-Smith, C. E., & Shippee, T. P. (2015). Expectations About Future Use Of Long-Term Services And Supports Vary By Current Living Arrangement. *Health Affairs*, *34*(1), 39–47. doi:10.1377/hlthaff.2014.0556

Herd, P., Goesling, B., & House, J. S. (2007). Socioeconomic Position and Health: The Differential Effects of Education versus Income on the Onset versus Progression of Health Problems. *Journal of Health and Social Behavior*, *48*(3), 223–238. doi:10.1177/002214650704800302

Human Fertility Database. (2015). *Completed cohort fertility*. Rostock, Germany. Retrieved from www.humanfertility.org

Hurd, M. D., Martorell, P., Delavande, A., Mullen, K. J., & Langa, K. M. (2013). Monetary Costs of Dementia in the United States. *New England Journal of Medicine*, *368*(14), 1326–1334. doi:10.1056/NEJMsa1204629

Imai, K., & Soneji, S. (2007). On the Estimation of Disability-Free Life Expectancy. *Journal of the American Statistical Association*, *102*(480), 1199–1211. doi:10.1198/016214507000000040

James, B. D., Leurgans, S. E., Hebert, L. E., Scherr, P. a, Yaffe, K., & Bennett, D. a. (2014). Contribution of Alzheimer disease to mortality in the United States. *Neurology*, *82*(12), 1045–50. doi:10.1212/WNL.0000000000000240

Johnson, E., Brookmeyer, R., & Ziegler-Graham, K. (2007). Modeling the Effect of Alzheimer's Disease on Mortality. *The International Journal of Biostatistics*, *3*(1). doi:10.2202/1557-4679.1083

Kalyani, R. R., Saudek, C. D., Brancati, F. L., & Selvin, E. (2010). Association of diabetes, comorbidities, and A1C with functional disability in older adults: results from the National Health and Nutrition Examination Survey (NHANES), 1999-2006. *Diabetes Care*, *33*(5), 1055–60. doi:10.2337/dc09-1597

Kasper, J. D., Freedman, V. A., Spillman, B. C., & Wolff, J. L. (2015). The Disproportionate Impact Of Dementia On Family And Unpaid Caregiving To Older Adults. *Health Affairs*, *34*(10), 1642–1649. doi:10.1377/hlthaff.2015.0536

Katz, M. J., Lipton, R. B., Hall, C. B., Zimmerman, M. E., Sanders, A. E., Verghese, J., … Derby, C. A. (2012). Age-specific and Sex-specific Prevalence and Incidence of Mild Cognitive Impairment, Dementia, and Alzheimer Dementia in Blacks and Whites. *Alzheimer Disease & Associated Disorders*, *26*(4), 335–343. doi:10.1097/WAD.0b013e31823dbcfc

Kemper, P., Komisar, H. L., & Alecxih, L. (2005). Long-term care over an uncertain future: What can current retirees expect? *Inquiry*, *42*(4), 335–350. doi:10.5034/inquiryjrnl_42.4.335

Khan, N. A., Wang, H., Anand, S., Jin, Y., Campbell, N. R., Pilote, L., & Quan, H. (2011). Ethnicity and sex affect diabetes incidence and outcomes. *Diabetes Care*, *34*(1), 96–101. doi:10.2337/dc10-0865

Lachin, J. M. (2011). *Biostatistical Methods*. *Biostatistical Methods: The Assessment of Relative Risks* (2nd ed.). Hoboken, NJ, USA: Wiley. doi:10.1002/9780470317051

Laditka, J., & Laditka, S. (2006). Effects of Diabetes on Healthy Life Expectancy: Shorter Lives with More Disability for Both Women and Men. In Z. Yi, E. Crimmins, Y. Carrière, & J.-M. Robine (Eds.), *Longer Life and Healthy Aging* (Vol. 2, pp. 71–90). Springer Netherlands. doi:10.1007/1-4020-4032-6_6

Langa, K. M., Larson, E. B., Karlawish, J. H., Cutler, D. M., Kabeto, M. U., Kim, S. Y., & Rosen, A. B. (2008). Trends in the prevalence and mortality of cognitive impairment in the United States: Is there evidence of a compression of cognitive morbidity? *Alzheimer's & Dementia*, *4*(2), 134–144. doi:10.1016/j.jalz.2008.01.001

Langa, Plassman, B. L., Wallace, R. B., Herzog, A. R., Heeringa, S. G., Ofstedal, M. B., … Willis, R. J. (2005). The Aging, Demographics, and Memory Study: study design and methods. *Neuroepidemiology*, *25*(4), 11.

Lee, H.-Y., Choi, C. S., Birkenfeld, A. L., Alves, T. C., Jornayvaz, F. R., Jurczak, M. J., … Shulman, G. I. (2010). Targeted Expression of Catalase to Mitochondria Prevents Age-Associated Reductions in Mitochondrial Function and Insulin Resistance. *Cell Metabolism*, *12*(6), 668–674. doi:http://dx.doi.org/10.1016/j.cmet.2010.11.004

Lee, J. M., Pilli, S., Gebremariam, A., Keirns, C. C., Davis, M. M., Vijan, S., … Gurney, J. G. (2010). Getting heavier, younger: trajectories of obesity over the life course. *Int J Obes*, *34*(4), 614–623. doi:10.1038/ijo.2009.235

Lee, P. G., Cigolle, C. T., Ha, J., Min, L., Murphy, S. L., Blaum, C. S., & Herman, W. H. (2013). Physical Function Limitations Among Middle-Aged and Older Adults With Prediabetes: One exercise prescription may not fit all. *Diabetes Care*, *36*(10), 3076–83. doi:10.2337/dc13-0412

Leffondré, K., Touraine, C., Helmer, C., & Joly, P. (2013). Interval-censored time-to-event and competing risk with death: Is the illness-death model more accurate than the cox model? *International Journal of Epidemiology*, *42*(4), 1177–1186. doi:10.1093/ije/dyt126

Li, N., Lee, R., & Gerland, P. (2013). Extending the Lee-carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography*, *50*(6), 2037–51. doi:10.1007/s13524-013-0232-2

Lönnroos, E., Kyyrönen, P., Bell, J. S., van der Cammen, T. J. M., & Hartikainen, S. (2013). Risk of death among persons with Alzheimer's disease: a national register-based nested case-control study. *Journal of Alzheimer's Disease*, *33*(1), 157–64. doi:10.3233/JAD-2012-120808

Loomis, D. (2005). Poisson regression analysis of ungrouped data. *Occupational and Environmental Medicine*, *62*(5), 325–329. doi:10.1136/oem.2004.017459

Loukine, L., Waters, C., Choi, B. C., & Ellison, J. (2012). Impact of diabetes mellitus on life expectancy and health-adjusted life expectancy in Canada. *Population Health Metrics*, *10*(1), 7. doi:10.1186/1478-7954-10-7

Lu, B. (2005). Propensity score matching with time-dependent covariates. *Biometrics*, *61*(3), 721–8. doi:10.1111/j.1541-0420.2005.00356.x

Luo, J., Rossouw, J., Tong, E., Giovino, G. A., Lee, C. C., Chen, C., … Margolis, K. L. (2013). Smoking and Diabetes: Does the Increased Risk Ever Go Away? *American Journal of Epidemiology*, *178*(6), 937–945. doi:10.1093/aje/kwt071

Manton, K. G., & Ukraintseva, S. V. (2005). Decline Prevalence of Dementia in the U.S . Elderly Population. *Advances in Gerontology*, *16*(73), 30–37.

Mason, W. M., & Fienberg, S. E. (1985). *Cohort Analysis in Social Research*. New York, NY: Springer-Verlag.

Mathers, C. D., & Robine, J. M. (1997). How good is Sullivan's method for monitoring changes in population health expectancies? *Journal of Epidemiology & Community Health*, *51*(1), 80–86. doi:10.1136/jech.51.1.80

McDermott, R. A., Li, M., & Campbell, S. K. (2010). Incidence of type 2 diabetes in two indigenous Australian populations: a 6-year follow-up study. *Medical Journal of Australia*, *192*(10), 4.

Meller, I., Fichter, M. M., & Schroppel, H. (1999). Mortality risk in the octo- and nonagenerians: longitudinal results of an epidemiological follow-up community study. *European archives of psychiatry and clinical neuroscience*, *249*(4), 180–189. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=10449593

Menke, A., Casagrande, S., Geiss, L., & Cowie, C. C. (2015). Prevalence of and Trends in Diabetes Among Adults in the United States, 1988-2012. *JAMA*, *314*(10), 1021. doi:10.1001/jama.2015.10029

Mingrone, G., Panunzi, S., De Gaetano, A., Guidone, C., Iaconelli, A., Leccesi, L., … Rubino, F. (2012). Bariatric surgery versus conventional medical therapy for type 2 diabetes. *N Engl J Med*, *366*(17), 1577–1585. doi:10.1056/NEJMoa1200111

Monesi, L., Baviera, M., Marzona, I., Avanzini, F., Monesi, G., Nobili, A., … Fortino, I. (2012). Prevalence, incidence and mortality of diagnosed diabetes: evidence from an Italian population based study. *Diabetic Medicine*, *29*(3), 385–392.

Mooney, C. Z. (1997). Monte Carlo Simulation. In *Sage University Paper Series on Quantitative Applications in the Social Sciences*. Thousand Oaks, CA: SAGE Publications.

Murphy, S. L., Xu, J., & Kochanek, K. D. (2013). *Deaths: Final Data for 2010* (Vol. 61). Hyattsville, MD: National Center for Health Statistics.

Murray, C. J. L. (2013). The state of U.S. health, 1990-2010: burden of diseases, injuries, and risk factors. *JAMA : the Journal of the American Medical Association*, *310*(6), 591–608. doi:10.1001/jama.2013.13805

Nagi, S. Z. (1976). An epidemiology of disability among adults in the United States. *The Milbank Memorial Fund quarterly. Health and society*, *54*(4), 439–67. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/137366

Narayan, K. M. V., 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. doi:10.2337/dc06-2544

Narayan, K. M. V., Boyle, J. P., Thompson, T. J., Sorensen, S. W., & Williamson, D. F. (2003). Lifetime risk for diabetes mellitus in the United States. *JAMA : the journal of the American Medical Association*, *290*(14), 1884–90. doi:10.1001/jama.290.14.1884

Narayan, K. M. V., Williams, D. E., Gregg, E. W., & Cowie, C. C. (2011). *Diabetes Public Health*. New York, NY: Oxford.

National Health and Nutrition Examination Survey. (2006). *Analytic and reporting guidelines*. National Center for Health Statistics. Retrieved from http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/nhanes_analytic_guidelines_de

c_2005.pdf

National Health and Nutrition Examination Survey. (2012). Survey questionnaires, datasets, and related documentation. Retrieved July 1, 2013, from http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm

National Health Interview Survey (1986-2004). (2009). Linked Mortality Files, mortality follow-up through 2006: Matching Methodology, May 2009. Hyattsville, MD: National Center for Health Statistics,.

National Institute of Diabetes and Digestive and Kidney Diseases. (2011). Longitudinal Assessment of Bariatric Surgery (LABS). Retrieved from http://win.niddk.nih.gov/publications/labs.htm

Ostbye, T., Hill, G., & Steenhuis, R. (1999). Mortality in elderly Canadians with and without dementia: a 5-year follow-up. *Neurology*, *53*(3), 521–526.

Palloni, A., & Thomas, J. R. (2013). Estimation of Covariate Effects With Current Status Data and Differential Mortality. *Demography*, *50*(2), 521–544. doi:10.1007/s13524-012-0160-6

Pavkov, M. E., Hanson, R. L., Knowler, W. C., Bennett, P. H., Krakoff, J., & Nelson, R. G. (2007). Changing patterns of type 2 diabetes incidence among Pima Indians. *Diabetes Care*, *30*(7), 1758–1763. doi:10.2337/dc06-2010

Perry, G. S., Byers, T. E., Mokdad, A. H., Serdula, M. K., & Williamson, D. F. (1995). The Validity of Self-Reports of Past Body Weights by U.S. Adults. *Epidemiology*, *6*(1), 61–66. doi:10.2307/3702451

Plassman, B. L., Langa, K. M., Fisher, G. G., Weir, D. R., Ofstedal, M. B., Burke, J. R., … Wallace, R. B. (2007). Prevalence of Dementia in the United States: the Aging, Demographics, and Memory Study. *Neuroepidemiology*, *29*(1), 8.

Plassman, B. L., Langa, K. M., McCammon, R. J., Fisher, G. G., Potter, G. G., Burke, J. R., … Wallace, R. B. (2011). Incidence of dementia and cognitive impairment, not dementia in the United States. *Annals of Neurology*, *70*(3), 418–26. doi:10.1002/ana.22362

Preston, S. H., Elo, I. T., Rosenwaike, I., & Hill, M. (1996). African-American mortality at older ages: results of a matching study. *Demography*, *33*(2), 193–209. doi:10.2307/2061872

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography*. Oxford: Blackwell.

Preston, S. H., & Wang, H. (2006). Sex mortality differences in the United States: The role of cohort smoking patterns. *Demography*, *43*(4), 631–646.

Raciborski, R. (2011). Right-censored Poisson regression model. *Stata Journal*, *11*(1), 95–105. Retrieved from http://phlib.kku.ac.th/StataJournalOnline/sj11-1.pdf#page=103

RAND. (2011). Rand HRS Data Documentation, Version L. Retrieved from http://hrsonline.isr.umich.edu/modules/meta/rand/randhrsl/randhrsL.pdf

Reis, J. P., Hankinson, A. L., Loria, C. M., Lewis, C. E., Powell-Wiley, T., Wei, G. S., & Liu, K. (2013). Duration of abdominal obesity beginning in young adulthood and incident diabetes through middle age: the CARDIA study. *Diabetes Care*, *36*(5), 1241–7. doi:10.2337/dc12-1714

Reither, E. N., Hauser, R. M., & Yang, Y. (2009). Do birth cohorts matter? Age-period-cohort analyses of the obesity epidemic in the United States. *Soc Sci Med*, *69*(10), 1439–1448. doi:10.1016/j.socscimed.2009.08.040

Rejeski, W. J., Bray, G. A., Chen, S.-H., Clark, J. M., Evans, M., Hill, J. O., … Ip, E. H. (2014). Aging and Physical Function in Type 2 Diabetes: 8 Years of an Intensive Lifestyle Intervention. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*. doi:10.1093/gerona/glu083

Rocca, W. A., Petersen, R. C., Knopman, D. S., Hebert, L. E., Evans, D. A., Hall, K. S., … White, L. R. (2011). Trends in the incidence and prevalence of Alzheimer's disease, dementia, and cognitive impairment in the United States. *Alzheimer's & Dementia*, *7*(1), 80–93. doi:10.1016/j.jalz.2010.11.002

Rosenbaum, P. R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika*, *74*(1), 13–26. Retrieved from http://biomet.oxfordjournals.org/content/74/1/13.short

Rosenbaum, P. R. (2002). *Observational Studies* (2nd ed.). New York, NY: Springer.

Rosenbaum, P. R. (2003). Exact Confidence Intervals for Nonconstant Effects by Inverting the Signed Rank Test. *The American Statistician*, *57*(2), 132–138. doi:10.1198/0003130031405

Rosenbaum, P. R. (2010). *Design of Observational Studies*. New York, NY: Springer.

Rosenbaum, P. R., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. Retrieved from http://biomet.oxfordjournals.org/content/70/1/41.short

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician*, *39*(1), 33–38.

Ruitenberg, A., Ott, A., & Swieten, J. van. (2001). Incidence of dementia: does gender make a difference? *Neurobiology of Aging*, *22*(June), 575–580. Retrieved from http://www.sciencedirect.com/science/article/pii/S0197458001002317

Salomon, J., Mathers, C., Murray, C., & Ferguson, B. (2001). *Methods for life expectancy and healthy life expectancy uncertainty analysis* (No. 10). Retrieved from http://cdrwww.who.int/entity/healthinfo/paper10.pdf

Schauer, P. R., Kashyap, S. R., Wolski, K., Brethauer, S. A., Kirwan, J. P., Pothier, C. E., … Bhatt, D. L. (2012). Bariatric surgery versus intensive medical therapy in obese patients with diabetes. *N Engl J Med*, *366*(17), 1567–1576. doi:10.1056/NEJMoa1200225

Schienkiewitz, A., Schulze, M. B., Hoffmann, K., Kroke, A., & Boeing, H. (2006). Body mass index history and risk of type 2 diabetes: results from the European Prospective Investigation into Cancer and Nutrition (EPIC)–Potsdam Study. *The American journal of clinical nutrition*, *84*(2), 427–433.

Selvin, E., Steffes, M. W., Zhu, H., Matsushita, K., Wagenknecht, L., Pankow, J., … Brancati, F. L. (2010). Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. *N Engl J Med*, *362*(9), 800–811. doi:10.1056/NEJMoa0908359

Seshadri, S., & Wolf, P. A. (2007). Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. *Lancet Neurology*, *6*(December), 1106–1114. doi:10.1016/S1474-4422(07)70291-0

Seshadri, S., Wolf, P. A., Beiser, A., Au, R., McNulty, K., White, R., & D'Agostino Sr., R. B. (1997). Lifetime risk of dementia and Alzheimer's disease: the impact of mortality on risk estimates in the Framingham Study. *Neurology*, *49*, 1498–1504. Retrieved from http://www.neurology.org/cgi/content/abstract/49/6/1498

Sheffield, K. M., & Peek, M. K. (2011). Changes in the Prevalence of Cognitive Impairment Among Older Americans, 1993-2004: Overall Trends and Differences by Race/Ethnicity. *American Journal of Epidemiology*, *174*(3), 274–283. doi:10.1093/aje/kwr074

Silber, J. H., Lorch, S. A., Rosenbaum, P. R., Medoff-Cooper, B., Bakewell-Sachs, S., Millman, A., … Escobar, G. J. (2009). Time to send the preemie home? Additional maturity at discharge and subsequent health care costs and outcomes. *Health Services Research*, *44*(2 Pt 1), 444–63. doi:10.1111/j.1475-6773.2008.00938.x

Silber, J. H., Rosenbaum, P. R., Trudeau, M. E., Even-Shoshan, O., Zhang, X., & Mosher, R. E. (2001). Multivariate and Bias Reduction in the Surgical Matching Outcomes Study. *Medical Care*, *39*(10), 1048–1064.

Singer, J., & Willett, J. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, *18*(2), 155–195. Retrieved from http://jeb.sagepub.com/content/18/2/155.short

Spillman, B. C., & Lubitz, J. (2002). New estimates of lifetime nursing home use: Have patterns of use changed? *Medical Care*, *40*(10), 965–975. doi:10.1097/01.MLR.0000027365.40102.7A

Stenholm, S., Westerlund, H., Head, J., Hyde, M., Kawachi, I., Pentti, J., … Vahtera, J. (2014). Comorbidity and Functional Trajectories From Midlife to Old Age: The Health and Retirement Study. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, (13), 1–7. doi:10.1093/gerona/glu113

Stokes, A., & Mehta, N. K. (2013). Mortality and excess risk in US adults with pre-diabetes and diabetes: a comparison of two nationally representative cohorts, 1988-2006. *Popul Health Metr*, *11*(1), 7. doi:10.1186/1478-7954-11-3

Sullivan, D. F. (1971). A single index of mortality and morbidity. *HSMHA health reports*, *86*(4), 347–54. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1937122&tool=pmcentrez&rendertype=abstract

Sylvestre, M., Huszti, E., & Hanley, J. (2006). Do Oscar winners live longer than less successful peers? A reanalysis of the evidence. *Annals of Internal Medicine*, *145*, 361–363. Retrieved from http://annals.org/article.aspx?articleid=728107&issueno=5&atab=10

The, N. S., Richardson, A. S., & Gordon-Larsen, P. (2013). Timing and duration of obesity in relation to diabetes: findings from an ethnically diverse, nationally representative sample. *Diabetes Care*, *36*(4), 865–872. doi:10.2337/dc12-0536

Thunander, M., Petersson, C., Jonzon, K., Fornander, J., Ossiansson, B., Torn, C., … Landin-Olsson, M. (2008). Incidence of type 1 and type 2 diabetes in adults and children in Kronoberg, Sweden. *Diabetes Res Clin Pract*, *82*(2), 247–255. doi:http://dx.doi.org/10.1016/j.diabres.2008.07.022

Tom, S. E., Hubbard, R. a, Crane, P. K., Haneuse, S. J., Bowen, J., McCormick, W. C., … Larson, E. B. (2015). Characterization of Dementia and Alzheimer's Disease in an Older Population: Updated Incidence and Life Expectancy With and Without Dementia. *American Journal of Public Health*, *105*(2), 408–413. doi:10.2105/AJPH.2014.301935

Treves, T. A., & Korczyn, A. D. (2012). Modeling the Dementia Epidemic. *CNS Neuroscience and Therapeutics*, *18*, 175–181. doi:10.1111/j.1755-5949.2011.00242.x

Tschanz, J., Corcoran, C., & Skoog, I. (2004). Dementia: The leading predictor of death in a defined elderly population The Cache County Study. *Neurology*, *62*, 1156–1162. Retrieved from http://www.neurology.org/content/62/7/1156.short

U.S. Census Bureau. (2005). Statistical Abstract of the United States, 2004-2005. Retrieved from http://www.census.gov/prod/2004pubs/04statab/pop.pdf

Vaupel, J. W. (2010). Biodemography of human ageing. *Nature*, *464*(7288), 536–42. doi:10.1038/nature08984

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, *16*(3), 439–454. doi:10.2307/2061224

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth Edi.). New York, NY: Springer.

Verbrugge, L. M., & Jette, A. M. (1994). The disablement process. *Social Science & Medicine*, *38*, 1–14. doi:10.1016/0277-9536(94)90294-1

Villarejo, A., Benito-León, J., Trincado, R., Posada, I. J., Puertas-Martín, V., Boix, R., … Bermejo-Pareja, F. (2011). Dementia-associated mortality at thirteen years in the NEDICES Cohort Study. *Journal of Alzheimer's Disease*, *26*(3), 543–51. doi:10.3233/JAD-2011-110443

Wilson, P. W. F., Anderson, K. M., & Kannel, W. B. (1986). Epidemiology of diabetes mellitus in the elderly: The Framingham study. *The American Journal of Medicine*, *80*(5, Supplement 1), 3–9. doi:http://dx.doi.org/10.1016/0002-9343(86)90532-2

Winkelmann, R. (2003). *Econometric Analysis of Count Data* (4th ed.). Berlin: Springer-Verlag.

Witthaus, E., Ott, A., Barendregt, J. J., Breteler, M., & Bonneux, L. (1999). Burden of mortality and morbidity from dementia. *Alzheimer Disease and Associated Disorders*, *13*(3), 176–181.

Wong, E., Backholer, K., Gearon, E., Harding, J., Freak-Poli, R., Stevenson, C., & Peeters, A. (2013). Diabetes and risk of physical disability in adults: a systematic review and meta-analysis. *The Lancet Diabetes & Endocrinology*, *1*(2), 106–114. doi:10.1016/S2213-8587(13)70046-9

Wong, & Gregg, E. W. (2013). Diabetes and risk of physical disability in adults: a systematic review and meta-analysis. *The Lancet Diabetes & Endocrinology*, *1*(2), 81–82. doi:10.1016/S2213-8587(13)70046-9

Wray, L. A., Ofstedal, M. B., Langa, K. M., & Blaum, C. S. (2005). The effect of diabetes on disability in middle-aged and older adults. *The journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *60*(9), 1206–11. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16183964

Yeh, H.-C., Duncan, B. B., Schmidt, M. I., Wang, N.-Y., & Brancati, F. L. (2010). Smoking, Smoking Cessation, and Risk for Type 2 Diabetes MellitusA Cohort Study. *Annals of Internal Medicine*, *152*(1), 10–17.

Zhao, W., Katzmarzyk, P. T., Horswell, R., Wang, Y., Li, W., Johnson, J., … Hu, G. (2014). Body Mass Index and the Risk of All-Cause Mortality Among Patients with Type 2 Diabetes. *Circulation*. doi:10.1161/CIRCULATIONAHA.114.009098

Ziegler-Graham, K., Brookmeyer, R., Johnson, E., & Arrighi, H. M. (2008). Worldwide variation in the doubling time of Alzheimer's disease incidence rates. *Alzheimer's & Dementia*, *4*(5), 316–323. doi:10.1016/j.jalz.2008.05.2479

Zissimopoulos, J., Crimmins, E., & St.Clair, P. (2014). The Value of Delaying Alzheimer's Disease Onset. *Forum for Health Economics and Policy*. doi:10.1515/fhep-2014-0013