Published in final edited form as:

J Stat Comput Simul. 2011 August ; 81(8): 973–983. doi:10.1080/00949650903583496.

Estimating age-specific incidence of dementia using prevalent cohort data

Binbing Yu^{a,*}

^aLaboratory of Epidemiology, Demography and Biometry, National Institute on Aging, Bethesda, MD 20892, USA

Abstract

In prospective cohort studies individuals are usually recruited according to a certain cross-sectional sampling criterion. The prevalent cohort is defined as a group of individuals who are alive but possibly with disease at the beginning of the study. It is appealing to incorporate the prevalent cases to estimate the incidence rate of disease before the enrollment. The method of back calculation of incidence rate has been used to estimate the incubation time from HIV infection to AIDS. The time origin is defined as the time of HIV infection. In aging cohort studies, the primary time scale is age of disease onset, subjects have to survive certain years to be enrolled into the study, thus creating left truncation (delay entry). The current methods usually assume that either the disease incidence is rare or the excess mortality due to disease is small compared to the healthy subjects. By far the validity of the results based on these assumptions has not been examined. In this paper, a simple alternative method is proposed to estimate dementia incidence rate before enrollment using prevalent cohort data with left truncation. Furthermore simulations are used to examine the performance of the estimation of disease incidence under different assumptions of disease incidence rates and excess mortality hazards due to disease. As application, the method is applied to the prevalent cases of dementia from the Honolulu Asia Aging Study to estimate dementia incidence rate and to assess the effect of hypertension, Apoe 4 and education on dementia onset.

Keywords

Dementia incidence; prevalent cohort; penalized likelihood

1. Introduction

The epidemiological concept of incidence is a rate of occurrence and is therefore connected to longitudinal follow-up of a population, while prevalence is the frequency in the population at a given point in time and thus relates to a cross-section sample. Usually both incidence and prevalence are required as components in the epidemiological characterization of a disease or other state, but sometimes only one of the two types of source of evidence is available. In many situations cross-sectional prevalent cohort studies are conducted because it does not require expensive follow-up of the cohort, especially for rare diseases.

The design of prevalent cohort studies is commonly used in epidemiologic studies, e.g., studies of infectious disease, AIDS and dementia [1]. A prevalent cohort is defined as a group of individuals who are alive but have a certain disease at study enrollment [2]. An

example of prevalent cohort concerns the estimation of age-specific incidence rates of dementia. The individuals may be alive but demented when enrolled. These cases of dementia are called prevalent cases. The individuals are then followed to identify the onset of new dementia cases, which are called incident cases. The main advantage of the prevalent cohort study is that it can be performed rapidly than the traditional cohort studies of individuals.

Figure 1 provides a schematic of a prevalent cohort study with follow-up that include both prevalent and incident cases. The study enrollment and cutoff dates are represented by two vertical lines. All individuals are alive in the beginning of the study. Cases 1 and 2 die without dementia and, thus, are treated as censored if dementia onset is considered as event. Cases 3 and 4 are prevalent cases as their dementia onset times are before study enrollment. Cases 5 and 6 are incident cases as the diagnosis of dementia occur within the active study period. In the usual analysis, the prevalent cases are removed as they do not contribute to the risk set of developing dementia [3]. Furthermore, the incidence cases can only contribute information of incidence rate after enrollment. It is natural to ask whether the additional information in the prevalent cases can be used to estimate the disease incidence (hazard) rates before enrollment.

The back-calculation method has been used to reconstruct past incidence rate of human immunodeficiency virus (HIV) infection based on prevalent cohort data [4]. This method either assumes a known distribution for the incubation period from HIV infection to AIDS onset [5,6] or known HIV infection times. Additional information can be incorporated into the analysis of the prevalent cohort. For example, in studies of dementia, Keiding [7,8] showed the inter-relationship between prevalence, incidence and survival. He also discussed how to use prevalent cohort data to estimate age-specific disease incidence rate under different scenarios. Again, the disease onset times for the prevalent cases are assumed to be known in the application.

In this article, a simple approximation method of estimating age-specific incidence rate with unknown disease onset times is proposed for prevalent cohort data with delay entry (left truncation). When the disease onset times are unknown, the proposed method and previous approaches [1,7] rely on the assumptions of rare disease or negligible excess mortality due to disease. Simulation studies are carried out to examine the performance of the proposed method under different assumptions about baseline incidence rate and excess mortality. The method is applied to the prevalent cases of dementia from Honolulu Asia Aging Study.

The rest of the paper is organized as follows: In Section 2, the study data is introduced and the major covariates are described. In Section 3, the statistical model for prevalent cohort data with unknown dementia onset times is described and the approximation of likelihood is derived. Then simulation results are shown in Section 4 and the proposed method is applied to the HAAS data in Section 5. The final section includes discussion and future research plans.

2. Honolulu Asia Aging Study

The Honolulu Heart Program (HHP) is a prospective study of heart disease and stroke for the cohort of 8006 Japanese-American men born between 1900 and 1919 who were living on the island of Oahu, Hawaii, at the time of baseline examination during the years 1965-1968 [9]. The average age of the cohort at baseline was 55 years old with minimum age 46 years and 95% percentile 65 years.

Three midlife examinations, conducted as part of the HHP in 1965, 1968–1970, and 1971-1974, included collection of clinical and demographic information. In 1991, the

Honolulu-Asia Aging Study (HAAS) was started as an extension of the HHP to study dementia prevalence, incidence and risk factors. At this examination (examination 4), 3,734 members of the HHP cohort participated. In examination 4, cognitive function and dementia were assessed by a multi-step procedure. The 100-point Cognitive Abilities Screening Instrument (CASI) was used to screen the entire sample [10]. A consensus diagnosis was reached by the study neurologist and at least two other physicians with expertise in geriatrics and dementia. In examination 4, 226 people were diagnosed with dementia and these cases were prevalent cases. White et al. [11] estimated the age-specific prevalence of dementia of the HAAS cohort, but did not examine the age-specific incidence rate.

In the subsequent three follow-up examinations conducted in 1994, 1997 and 2000, cognitive function and dementia were continually assessed. The cases diagnosed in examinations 5–7 are incident cases. In the previous analysis of HAAS, only the incident cases were used and the prevalent cases were excluded [12,13]. Because the prevalent cases contained information about incidence rate before examination 4, it is appealing to use them to estimate the dementia onset rate for younger ages. Here the focus of the paper is on how to use the prevalent cases to estimate the incidence rate of dementia under reasonable assumptions, which is a simple alternative to the methods that use incident cases.

3. Statistical Method

A basic tool to model each individual's dynamics is a simple three state stochastic 'illness-death' process, which uses age as the time scale and interprets the age-specific disease rate as incidence. The *illness-death model* assumes that the subjects start out as healthy (State 1), and may become ill (State 2) and die (State 3) afterwards, and the subjects may die without having the disease (Figure 2). Let T denote the age of death. The age-specific incidence rate of disease is $\lambda_{12}(t)$, and the intensities of death for non-diseased and diseased subjects are $\lambda_{13}(t)$ and $\lambda_{23}(t)$, respectively. The intensity of mortality for diseased subjects could depend on disease onset age as well as covariates, leading to a semi-Markov model.

Let X be the vector of covariates, which could be time-dependent. We assume that the age-specific disease rate is given by

$$\lambda_{12}(t|X) = \lambda_{12,0}(t) \exp(\beta^T X), \quad (1)$$

where $\lambda_{12,0}(t)$ is the baseline hazard function. Various models can be used for the baseline hazard. For the Cox proportional hazards model, $\lambda_{12,0}(t)$ is non-parametric, and for parametric models, $\lambda_{12,0}(t)$ can be specified by a two- or three-parameter function. For example, for the Weibull model the hazard function is

$$W(t|a, b) = \frac{a}{b} \left(\frac{t}{b} \right)^{a-1},$$

where a is the shape parameter and b is the scale parameter. The estimation methods of the illness-death model for right censored data are described by Putter et al. [14] and Meira-Machado et al. [15]. For survival data with left truncation and interval censoring, a multiple imputation method can be used to estimate $(\lambda_{12,0}(t), \lambda_{13,0}(t), \lambda_{23,0}(t))$ for prevalent cohort survival data with information on death and disease onset. This method involves imputation of unknown disease status for the subjects who died before the study began and is computation intensive. Here the focus is the estimation of age-specific disease rate $\lambda_{12}(t)$ and its association with covariates. Thus, the parameters of interest are $\theta = (\lambda_{12,0}, \beta)$ and the other parameters and intensities are nuisance parameters.

Let $\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(s) ds$ and $S_{ij} = \exp\{-\Lambda_{ij}(t)\}$ for $1 \leq i < j \leq 3$ and let $S_1(t) = \exp\{-\Lambda_{12}(t) - \Lambda_{13}(t)\}$. We assume that all subjects start alive and dementia-free at examination 1 and let the age at examination 1 denoted by u . The analysis is based on the cohort at examination 4, which includes the prevalent cases of dementia. For a subject at examination 4, let v denote the current age and let d be the binary indicator for his disease status. If a subject is disease-free ($d = 0$), the likelihood contribution is

$$L_1 = S_1(v) / S_1(u).$$

If a subject has disease, the likelihood contribution is given by

$$L_2 = F(u, v) / S_1(u),$$

where

$$F(u, v) = \int_u^v S_1(r) \lambda_{12}(r) \exp\left(-\int_r^v \lambda_{23}(s) ds\right) dr$$

is the probability of developing disease during $[u, v]$ and being alive at v .

Without loss of generality, we assume that the first n_1 subjects are alive and disease free and the next n_2 subjects are alive with disease at the enrollment of the study, which is examination 4 in the HAAS. The total number of subjects is $n = n_1 + n_2$. The likelihood function for the observed data is

$$L(\theta) = \prod_{i=1}^{n_1} \frac{S_1(v_i)}{S_1(u_i)} \prod_{i=n_1+1}^n \frac{F(u_i, v_i)}{S_1(u_i)}, \quad (2)$$

where (u_i, v_i) are the ages of the i th subject at examinations 1 and 4, respectively. The likelihood function $L(\theta)$ involves the death intensities $\lambda_{13}(t)$ and $\lambda_{23}(t)$. The parameters can be obtained by maximizing likelihood function (2). When the retrospective information on death prior to the prevalent cohort study was not available, the parameter estimates can be obtained by approximating the likelihood (2). When the excess mortality due to dementia is small, i.e., $\lambda_{23}(t) \approx \lambda_{13}(t)$, then $F(u, v)$ can be approximated as

$$\begin{aligned} F(u, v) &= \int_u^v \exp\left\{-\int_0^r \lambda_{12}(s) ds\right\} \lambda_{12}(r) \times \exp\left\{-\int_0^r \lambda_{13}(s) ds - \int_r^v \lambda_{23}(s) ds\right\} dr \\ &\approx \int_u^v \exp\left\{-\int_0^r \lambda_{12}(s) ds\right\} \lambda_{12}(r) \times \exp\left\{-\int_0^r \lambda_{13}(s) ds\right\} dr \\ &= S_{13}(v) \{S_{12}(u) - S_{12}(v)\}. \end{aligned} \quad (3)$$

Thus, the likelihood function $L(\theta)$ can be written as

$$L(\theta) = C \prod_{i=1}^{n_1} \frac{S_{12}(v_i)}{S_{12}(u_i)} \prod_{i=n_1+1}^n \frac{S_{12}(u_i) - S_{12}(v_i)}{S_{12}(u_i)}, \quad (4)$$

where $C = \prod_{i=1}^n \{S_{13}(v_i) / S_{13}(u_i)\}$ only involves the death intensity $\lambda_{13}(t)$. The formulation (4) is in fact the likelihood function for survival data with interval censoring and left truncation. Joly et al. [16] propose a penalized-likelihood method to estimate the baseline hazard function. They first introduce an *a priori* knowledge of smoothness of the hazard

function by penalizing the likelihood function. The smoothed hazard function is approximated on a basis of splines and is obtained by maximizing the penalized likelihood. A user-friendly Fortran program PHMPL is also provided by Joly et al. [3], which can be used to find the maximum likelihood estimates of $(\lambda_{12,0}, \beta)$ based on likelihood (4).

4. Simulation Study

Simulations are carried out to examine the performance of the proposed estimator, especially when the assumptions of rare disease and small excess mortality do not hold. We also examine the effect under different model specifications, i.e., additive hazards and proportional hazards model.

To mimic the real studies of dementia, we assume every subject is dementia-free with minimum age 45. So the actual onset time variable T is dementia onset age minus 45. We assume that only one factor X is related to disease and X is binary with $P(X = 1) = 0.2$. The disease incidence rate is

$$\lambda_{12}(t) = \lambda_{12,0}(t) \exp(\beta X),$$

where we set $\beta = 0.5$ and the baseline hazard $\lambda_{12,0}(t) = W(t|a_1, b_1) \exp(\nu)$ with $a_1 = 6$ and $b_1 = 55$. The proportionality factor $\nu = -0.5$ or 0.5 to create moderate and relatively high age-specific disease rates. Figure 3 shows the baseline disease incidence rate corresponds to the two ν values. The disease incidence rate is very low before age 65 for both incidence groups. The low incidence group reflects the age-specific incidence rates of cancer, which are below 4% for males and below 2% for females for age greater than 75 [17]. The high incidence group reflects the age-specific rates of dementia, which are about 7% for males and 5% for females at age 90 based on PAQUID study [18].

The hazard of death for healthy subjects is $W(t|a_2, b_2)$ with $a_2 = 3.5$ and $b_2 = 40$. Two possible models are used for the excess mortality of diseased subjects. For additive excess mortality, $\lambda_{23}(t) = \lambda_{13}(t) + \delta$ and for multiplicative excess mortality $\lambda_{23}(t) = \lambda_{13}(t) \exp(\delta)$. The possible values for δ are shown in Table 1. Figure 4 plots the baseline mortality rate $\lambda_{13}(t)$ and excess mortality rate $\lambda_{23}(t)$ under different scenarios.

In the prevalent cohort studies, a common problem is delay entry because people have to survive to certain ages to be enrolled into the study. Here the entry age follows a Weibull distribution $W(t|a_0, b_0)$ with $a_0 = 1.87$ and $b_0 = 11.2$. The median entry age is $5.6 + 45 = 50.6$ years. The actual history of the subjects is simulated according to the illness-death model. Because we assume that all subjects are dementia-free in the beginning of the study, the subjects who die or develop dementia before the enrollment are excluded. The time from disease onset to death for demented subject is simulated according to the algorithm described in Appendix A. The results are based on 1000 simulations and in each simulation 8000 subjects are generated.

The estimates of $(\beta, \lambda_{12,0}(t))$ is denoted by $(\widehat{\beta}, \widehat{\lambda_{12,0}}(t))$. To summarize our simulation results, we compute the average bias of the estimate $\widehat{\beta}$ and actual coverage rate (CR) of the 95% confidence interval and we compare the mean of the estimated standard deviation (MSD) and the empirical standard deviation (ESD) of $\widehat{\beta}$. We also calculate the bias of the smoothed incidence rate curve. The average bias and average absolute bias of the incidence rate curve are calculated based on 50 equidistant points, $t_{(1)}, \dots, t_{(50)}$, on the age grid from 45 to 90. For example, the average bias of $\widehat{\lambda_{12,0}}(t)$ for age between 45 and 90 is calculated as

$$\text{Bias}(\widehat{\lambda_{12,0}}) = \text{Mean of } \left\{ \widehat{\lambda_{12,0}}(t_{(i)}) - \lambda_{12,0}(t_{(i)}) : t_{(i)} \in [45, 90] \right\}.$$

Table 2 shows the average bias of $\widehat{\beta}$ and the MSD and ESD. We see that the biases of $\widehat{\beta}$ from the multiplicative model are small, while the biases from the additive model are larger, especially for models with low baseline disease rate and high excess mortality. For the additive model with high excess mortality, more subjects in the $X = 1$ group tend to die earlier than those in the $X = 0$ group and these subjects are not observed because of delay entry. Thus, the coefficient of X is under estimated based on observed subjects with disease. The actual coverage rates of the 95% confidence interval are higher than the nominal level, especially for disease with lower incidence rate. The MSDs are larger than the ESDs. Thus the confidence interval based on the estimated standard deviation has a wider range than it should be and the corresponding Wald tests are conservative. Overall, we see that the parameter estimate $\widehat{\beta}$ has small bias for models except for additive model with excess mortality parameter $g=d \leq 0.10$.

Table 3 shows the bias and absolute bias of disease incidence rates expressed in percentage. The incidence rate for age less than 65 is close to 0 and the estimated incidence curve has very small bias. For age between 65 and 90, the average absolute incidence rates are under 1%. The biases are more negative for disease with higher excess mortality. Based on the simulation, one needs to be cautious to use the prevalent cohort data to estimate the disease incidence rate when the excess mortality rate is high and the disease is not rare.

5. Application

Because 95% of the subjects at the beginning of the HHP program are below 65 years old and the dementia incidence before age 65 is rare, we assume that the original cohort of 8006 people are dementia-free. We estimate the dementia incidence rate based on the prevalent cohort of 3734 people at examination 4 and assess the effect of three major risk factors. The three covariates used in the analysis are education level (edu), late-life hypertension (hyp), and ApoE genotype (apoe). The variable for hypertension is the medical history of high blood pressure at examination 1, the education level is a binary variable to indicate whether the total number of years of education is more than 10 years, and ApoE genotype is a indicator of carrying $\epsilon 4$ allele. In the analysis cohort, 59% had education more than 10 years, 17% reported hypertension at exam 1, and 18% were ApoE $\epsilon 4$ carriers. The reference group are the subjects with lower education, no hypertension history, and without ApoE $\epsilon 4$ allele.

The proportional hazards assumption is examined by plotting the estimated hazard functions $\lambda_{13}(t)$ and $\lambda_{23}(t)$ in the illness-death model in Figure 2 without the restriction of proportionality. The assumption of proportional hazards is well satisfied. Thus, a proportional hazards model is used for the dementia incidence rates

$$\lambda_{12}(t|Z) = \lambda_{12,0}(t) \exp(\beta_1 \text{apoe} + \beta_2 \text{edu} + \beta_3 \text{hyp}).$$

The PHMPL program [3] is used to computer the parameters $(\beta_1, \beta_2, \beta_3)$ and the baseline dementia incidence rate $\lambda_{12,0}(t)$.

The estimated baseline dementia incidence curve $\widehat{\lambda_{12,0}}(t)$ is plotted in Figure 5. We see that the incidence rates are very close to 0 until age 65. Then the dementia incidence rates gradually increases. For the subjects at 85 years old, the annual incidence rate is about 3%, which is slightly lower than the rate of 3.5% based on the HAAS follow-up data. The overall

incidence curve is very similar to the estimated incidence curve for men based on the PAQUID study [18, p207]. The parameter estimates, standard error, p -value from the Wald test, hazard ratio and the 95% confidence interval are shown in Table 4. As we demonstrate by simulation, the Wald tests based on the estimated S.E. tend to be conservative. We see all the variables are highly significantly related to dementia onset. The carriers of ApoE $\epsilon 4$ genotype and subjects with hypertension have higher risk of developing dementia. The subjects with higher education have lower risk of dementia onset.

6. Discussion

In this article, we proposed a simple approximation method to estimate the disease incidence rate using prevalent cohort data. The performance of the estimates of baseline hazard $\lambda_{12,0}(t)$ and the regression coefficients β are examined under different assumptions of disease incidence and excess mortality rate by simulations. The simulation results can serve as a guide to use the prevalent cohort data to estimate disease incidence rate before the enrollment.

In particular, as prevalence is the product of incidence and duration, a covariate effect on duration may be biased due to biased estimates for incidence. While dementia is rare before the age of 65, it is not evident whether there is negligible excess mortality. As shown by the simulations, the approximation method can have issues when excess mortality is high. Some additional information is required for bias correction. If the excess mortality due to dementia is known from historic data and similar studies [19,20], one can plug the updated mortality rate $\lambda_{23}(t)$ in (3) to obtain unbiased estimates of covariate effects. If the additional information of time since diagnosis is available for each prevalent case, one can utilize such information to adjust for the length bias due to different survival. In the setting of logistic regression, Begg and Gray [21] proposed a method to correct the bias in the odds ratio estimates relating to the exposures to disease occurrence. This method can be extended to the adjust for the bias on hazard ratio of dementia incidence rate.

If the prevalent cohort is also followed up to ascertain disease onset or death, the follow-up data with incident cases can be combined with the prevalent cases to estimate the age-specific incidence rate. If we assume the disease incidence does not change, using both prevalent and incident cases can increase the power of the analysis. If the trend of disease incidence is of interest, one can compare the prevalent cohort and follow-up data to examine the trend of stationarity of the incidence rate [22]. A formal test for stationarity has also been proposed by Addona and Wolfson [23].

The proposed method is especially useful when only the cross-sectional data is available, as this is the only choice to estimate the disease incidence. The resulting estimates can be used in the design of the prospective follow-up study of the cohort. Given the advantages of time efficiency and larger sample, analysts should keep in mind the major problems, e.g., recall bias and length bias, of using prevalent cohort data.

Acknowledgments

The research was supported in part by the Intramural Research Program of the National Institute on Aging.

Appendix A. Simulation of the time T23 from dementia onset to death

Let T_{ij} denote the time from state i to state j . Then

$$S(t_{23}|t_{12}) = S(T_{23} > t_{23} | T_{12} = t_{12}) = \exp\left\{-\int_{t_{12}}^{t_{12}+t_{23}} \lambda_{23}(t) dt\right\}$$

For multiplicative model, $\lambda_{23}(t) = \lambda_{13}(t) \exp(\delta)$, then

$$S(t_{23}|t_{12}) = \exp\left\{-\int_{t_{12}}^{t_{12}+t_{23}} \lambda_{3,0}(t) dt\right\}^{\exp(\delta)} = \left(\frac{S_{3,0}(t_{12}+t_{23})}{S_{3,0}(t_{12})}\right)^{\exp(\delta)}$$

For the Weibull model, $S_{3,0}(t) = \exp\{-(t/b_3)^{a_3}\}$. To generate time T_{23} for demented people with dementia onset time t_{12} . First, a uniform random variable U is generated, then t_{23} is calculated as

$$t_{23} = b_3 \left[\left(\frac{t_{12}}{b_3} \right)^{a_3} - \frac{\log U}{\exp(\delta)} \right]^{\frac{1}{a_3}} - t_{12}.$$

For additive model, $\lambda_{23}(t) = \lambda_{13}(t) + \delta$, then

$$S(t_{23}|t_{12}) = \left(\frac{S_{3,0}(t_{12}+t_{23})}{S_{3,0}(t_{12})} \right) \exp(-\delta t_{23})$$

generate time T_{23} for demented people with dementia onset time t_{12} . First, two uniform random variables U_1 and U_2 are generated, then $t_{23}^{(1)}$ is calculated as

$$t_{23}^{(1)} = b_3 \left[\left(\frac{t_{12}}{b_3} \right)^{a_3} - \log U_1 \right]^{\frac{1}{a_3}} - t_{12}.$$

and $t_{23}^{(2)} = -(\log U_2) / \delta$. Then $t_{23} = \min(t_{23}^{(1)}, t_{23}^{(2)})$.

References

1. Shkedy Z, Aerts M, Molenberghs G, Beutels Ph, Van Damme P. Modelling age-dependent force of infection from prevalence data using fractional polynomials. *Statistics in Medicine*. 2006; 25:1577–1591. [PubMed: 16252265]
2. Wang M-C, Brookmeyer R, Jewell NP. Statistical models for prevalent cohort data. *Biometrics*. 1993; 49:1–11. [PubMed: 8513095]
3. Joly P, Letenneur L, Alioum A, Commenges D. PHMPL: a computer program for hazard estimation using a penalized likelihood method with interval-censored and left-truncated data. *Computer Methods and Programs in Biomedicine*. 1999; 60:225–231. [PubMed: 10579515]
4. Bacchetti P, Segal MR, Jewell NP. Backcalculation of HIV infection rates. *Statistical Sciences*. 1993; 2:82–101.
5. Brookmeyer R. Statistical methods for short-term projections of AIDS incidence. *Statistics in Medicine*. 1989; 8:23–34. [PubMed: 2919244]
6. Brookmeyer R, Gail MH. A method of obtaining short-term predictions and lower bounds on the size of the AIDS epidemic. *Journal of the American Statistical Association*. 1989; 83:301–308.
7. Keiding N. Age-specific incidence and prevalence: a statistical perspective. *Journal of the Royal Statistical Society, Series A*. 1991; 154:371–412.

8. Keiding N. Event history and the cross-section. *Statistics in Medicine*. 2006; 25:2343–2364. [PubMed: 16708345]
9. Heilbrun LK, Kagan A, Nomura A, Wasnich RD. The origins of epidemiologic heart diseases, cancer and Osteoporosis among Hawaii Japanese. *Hawaii Medical Journal*. 1985; 44:294–296. [PubMed: 4055356]
10. Teng EL, Hasegawa K, Homma A, Imai Y, Larson E, Graves A, Sugimoto K, Yamaguchi T, Sasaki H, Chiu D, et al. The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia. *International Psychogeriatrics*. 1994; 6:45–58. [PubMed: 8054493]
11. White L, Petrovitch H, Ross GW, Masaki KH, Abbott RD, Teng EL, Rodriguez BL, Blanchette PL, Havlik RJ, Wergowske G, Chiu D, Foley DJ, Murdaugh C, Curb JD. Prevalence of dementia in older Japanese-American men in Hawaii: The Honolulu-Asia Aging Study. *Journal of the American Medical Association*. 1996; 276:955–960. [PubMed: 8805729]
12. Havlik RJ, Izmirlian G, Petrovitch H, Ross GW, Masaki K, Curb JD, Saunders AM, Foley DJ, Brock D, Launer LJ, White L. APOE- ϵ 4 predicts incident AD in Japanese-American men: The Honolulu-Asia Aging Study. *Neurology*. 2000; 54:1526–1529. [PubMed: 10751272]
13. Saczynski JS, White L, Peila RL, Rodriguez BL, Launer LJ. The relation between apolipoprotein A-I and dementia: the Honolulu-Asia aging study. *American Journal of Epidemiology*. 2007; 165:985–992. [PubMed: 17298957]
14. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*. 2007; 26:2389–2430. [PubMed: 17031868]
15. Meira-Machado L, Cadarso-Suárez C, De Uña-Álvarez J. tdc.msm: An R library for the analysis of multi-state survival data. *Computer Methods and Programs in Biomedicine*. 2007; 86:131–140. [PubMed: 17350136]
16. Joly P, Commenges, Letenneur L. A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. *Biometrics*. 1998; 54:185–194. [PubMed: 9574965]
17. Merrill RM, Weed DL. Measuring the Public Health Burden of Cancer in the United States Through Lifetime and Age-Conditional Risk Estimates. *Annals of Epidemiology*. 2001; 11:547–553. [PubMed: 11709274]
18. Commenges D, Joly P, Letenneur L, Dartigues JF. Incidence and mortality of Alzheimer's disease or dementia using an illness-death model. *Statistics in Medicine*. 2004; 23:199–210. [PubMed: 14716722]
19. Ueki A, Sinjo H, Shimode H, Nakajima T, Morita Y. Factors associated with mortality in patients with early-onset Alzheimer's disease: a five-year longitudinal study. *International Journal of Geriatric Psychiatry*. 2001; 16:810–815. [PubMed: 11536348]
20. Dewey ME, Saz P. Dementia, cognitive impairment and mortality in persons aged 65 and over living in the community: a systematic review of the literature. *International Journal of Geriatric Psychiatry*. 2001; 16:751–761. [PubMed: 11536341]
21. Begg CB, Gray RJ. Methodology for case-control studies with prevalent cases. *Biometrika*. 1987; 74:191–195.
22. Asgharian M, Wolfson DB, Zhang X. Checking stationarity of the incidence rate using prevalent cohort survival data. *Statistics in Medicine*. 2006; 25:1751–1767. [PubMed: 16220462]
23. Addona V, Wolfson DB. A formal test for stationarity of the incidence rate using data from a prevalent cohort study with followup. *Lifetime Data Analysis*. 2006; 12:267–274. [PubMed: 16917734]

Figure 1.
A schematic of prevalent cohort data

Figure 2.
The illness-death model: λ_{12} , λ_{13} , and λ_{23} are the intensity functions

Figure 3.
Baseline disease incidence rate $\lambda_{12,0}(t)$

Figure 4.
Baseline mortality rate $\lambda_{13}(t)$ and mortality rate $\lambda_{23}(t)$ under different scenarios

Figure 5.
Dementia incidence rate estimated from prevalent cohort data

Table 1

Values of δ for excess mortality rate due to disease

Excess mortality	Model	
	Additive	Multiplicative
Small	0.02	0.25
Moderate	0.10	0.50
Large	0.15	1.00

Table 2

Bias, MSD and ESD of $\hat{\beta}$

Baseline rate	Model	δ	Bias	CR %	MSD	ESD
Low ($v = -0.5$)	Multiplicative	0.25	-0.018	98.3	0.157	0.147
		0.50	-0.021	97.8	0.164	0.151
		1.00	-0.019	98.0	0.196	0.168
	Additive	0.02	-0.030	98.2	0.220	0.147
		0.10	-0.103	98.7	0.231	0.192
		0.15	-0.128	99.4	0.228	0.207
High ($v = 0.5$)	Multiplicative	0.25	-0.002	98.3	0.173	0.093
		0.50	0.006	97.4	0.131	0.097
		1.00	0.030	95.5	0.152	0.111
	Additive	0.02	-0.006	98.6	0.185	0.095
		0.10	0.027	96.5	0.175	0.126
		0.15	0.049	94.4	0.153	0.144

Table 3

Bias and absolute bias of disease incidence rates by age expressed in percentage

Model	δ	v	≤ 65		[65 – 90]		[45 – 90]	
			Bias	Bias	Bias	Bias	Bias	Bias
Multiplicative	0.25	-0.5	0.01	0.03	-0.54	0.54	-0.31	0.32
	—	0.5	0.04	0.06	0.31	0.78	0.19	0.47
	0.50	-0.5	0.01	0.03	-0.61	0.61	-0.35	0.37
	—	0.5	0.04	0.06	0.05	0.80	0.05	0.49
	1.00	-0.5	0.01	0.03	-0.82	0.82	-0.47	0.49
Additive	—	0.5	0.04	0.06	-0.45	0.86	-0.24	0.52
	0.02	-0.5	0.01	0.03	-0.47	0.48	-0.26	0.29
	—	0.5	0.04	0.06	0.61	0.77	0.37	0.47
	0.10	-0.5	0.00	0.02	-0.95	0.97	-0.55	0.56
	—	0.5	0.01	0.01	-0.66	0.75	-0.38	0.44
	0.15	-0.5	0.00	0.02	-1.08	1.08	-0.62	0.63
		0.5	0.00	0.01	-0.88	0.92	-0.51	0.53

Table 4

Estimates of the hazard ratio (HR) on dementia for the risk factors

Variable	Estimate	S.E.	p-value	HR	95% CI
apoe	0.557	0.161	<.001	1.745	(1.273, 2.392)
hyp	0.523	0.157	<.001	1.687	(1.241, 2.293)
stdedu	-0.430	0.148	0.004	0.650	(0.486, 0.870)