Published in final edited form as:

Stat Med. 2019 May 30; 38(12): 2103–2114. doi:10.1002/sim.8098.

Analysis of combined incident and prevalent cohort data under a proportional mean residual life model

Chi Hyun Lee[*,1], **Jing Ning**[1], **Richard J. Kryscio**[2,3], and **Yu Shen**[1]

[1]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas, U.S.A.

[2]Department of Biostatistics, Sanders-Brown Center on Aging, University of Kentucky, Lexington, Kentucky, U.S.A.

[3]Department of Statistics, University of Kentucky, Lexington, Kentucky, U.S.A.

Summary

The Nun Study, a longitudinal study to examine risk factors for the progression of dementia, consists of subjects who were already diagnosed with dementia (i.e., prevalent cohort) and those who do not have dementia (i.e., incident cohort) at study enrollment. When assessing the risk factors' effects on the survival time from dementia diagnosis until death, utilizing data from both cohorts supports more efficient statistical inference because the two cohorts provide valuable complementary information. A major challenge in analyzing the combined cohort data is that the prevalent cases are not representative of the target population. Moreover, the dates of dementia diagnosis are not ascertained for the prevalent cohort in the Nun Study. Hence, the survival time for the prevalent cohort is only partially observed from study enrollment until death or censoring, with the time from dementia diagnosis to study enrollment missing. In this paper, we propose an efficient estimation method that uses both incident and prevalent cohorts under the proportional mean residual life model. By assuming proportionality of the mean residual life time with covariates in the incident cohort, we can utilize the natural relationship between the mean residual life function and the hazard function of the survival time measured from enrollment until death for the prevalent cohort. We evaluate the efficiency gain from using the combined cohort data through simulations and demonstrate that the proposed method is valid and efficient.

Keywords

Combined cohort data; incident cohort; Nun Study; prevalent cohort; proportional hazards model; proportional mean residual life model

1 | INTRODUCTION

Prospective observational studies are commonly used to identify and evaluate risk factors that are associated with disease-specific survival. Such studies occasionally include both incident and prevalent cohorts. For example, the Nun Study of Aging and Alzheimer's

[*]**Correspondence** Chi Hyun Lee, 1400 Pressler Street Unit 1411, Houston, Texas 77030, U.S.A. clee9@mdanderson.org.

Disease (Nun Study),[1] which motivates this work, involves an incident cohort of subjects who have not experienced dementia onset and are followed over time to monitor the potential diagnosis of dementia and death; and the prevalent cohort of subjects who already have dementia but have not experienced death at the time of study entry. The two cohorts provide valuable complementary information: the incident cohort is a random sample from the target population; and the prevalent cohort includes more deaths since subjects are sampled in the midst of dementia. Thus, analyzing the combined data from both cohorts yields more efficient statistical results. However, statistical analysis using the combined data has received less attention in the literature.

The data from the Nun Study consist of 501 subjects after excluding 177 participants who had missing key covariates (22) or withdrew consent (155). Among the participants represented in the data, 77 (about 15%) already had dementia and 424 were not yet diagnosed with dementia at study entry; these participants comprise the prevalent and incident cohorts, respectively. During the prospective follow-up, 153 subjects among the incident cohort were diagnosed with dementia. The dates of diagnosis of dementia were not available for the 77 subjects with dementia in the prevalent cohort. The combined cohort data are illustrated in Figure S1 of the web-based supplementary materials. In the statistical literature, the Nun Study data have been used primarily to illustrate Markov transition models,[2,3,4,5,6] which has excluded the data from the prevalent cohort. We aim to take advantage of data from both the prevalent and incident cohorts for more efficient evaluation of the relationship between the risk factors and the survival time after diagnosis of dementia. In addition to the challenge of properly adjusting for sampling bias, a major issue when analyzing the combined data from the Nun Study is that the dates of dementia diagnosis for the prevalent cases were not ascertained. Thus, we only observe the time from study enrollment to death (referred to as the “forward recurrence time”) with the information of the time from diagnosis of dementia to study enrollment (referred to as the “backward recurrence time”) missing for the prevalent cohort.

We consider the proportional mean residual life (PMRL) model [7] to assess the effect of risk factors on the residual survival time. By assuming proportionality of the mean residual life time with covariates, we can utilize the natural relationship between the mean residual life function and the hazard function of the forward recurrence time to analyze the combined cohort data. In Section 2, we introduce notations to depict the combined cohort data and present the connection between the PMRL model and the proportional hazards (PH) model. We review existing estimation methods for data from the incident cohort only and the prevalent cohort only, and propose efficient estimating equations for the combined cohorts in Section 3. The asymptotic properties are also established in this section. We investigate finite sample properties through simulation studies under various settings in Section 4. In Section 5, we use the proposed method to analyze the Nun Study data. We provide some remarks in Section 6.

2 | NOTATIONS AND MODEL

We consider data from both the incident and prevalent cohorts with respective sample sizes of n_1 and n_2. For the incident cohort, we denote T^0 and a $p \times 1$ vector $\boldsymbol{X}$ as the duration from disease diagnosis to death and the time-independent covariates, respectively. Let C be the duration from disease diagnosis to a censoring event. Then, the observed data from the incident cohort consist of independent and identically distributed (i.i.d.) $\{(T_i, \mathrm{A}_i, \boldsymbol{X}_i), i = 1, \ldots, n_1\}$, where $T_i = \min(T_i^0, C_i)$ and $\Delta_i = I(T_i^0 \le C_i)$. We assume that the censoring time $\boldsymbol{C}$ is conditionally independent of T^0 given covariates $\boldsymbol{X}$. We note that the incident cohort is representative of the target population. For prevalent cases, the dates of dementia diagnosis, which occurred prior to enrollment, are unknown. Thus, only partial information on survival times that is measured from the study enrollment is available. We introduce additional notations to represent the event times observed from the prevalent cohort. Let V^0 and a $p \times 1$ vector $\boldsymbol{X^v}$ denote the duration from enrollment to death and the time-independent covariates for the prevalent cohort, respectively. Unlike the censoring time C for the incident cohort, the censoring time C^v is measured from enrollment until a censoring event. The observed prevalent cohort data are i.i.d. $\{(V_i, \Delta_i^v, \boldsymbol{X}_i^v), i = 1, \ldots, n_2\}$, where $V_i = \min(V_i^0, C_i^v)$ and $\Delta_i^v = \mathrm{I}(V_i^0 \le C_i^v)$. The censoring time for the prevalent cohort, C^v, is assumed to be conditionally independent of $\boldsymbol{V}^0$ given covariates $\boldsymbol{X}^v$. Based on research about dementia,[8,9] it is reasonable to assume that the natural history of dementia follows a stationary Poisson process. Under such an assumption, the prevalent cohort is subject to length-biased sampling.

The mean residual life function for the underlying survival time T^0 at time t can be defined as $m(t \mid \boldsymbol{X}) = \mathrm{E}(T^0 - t \mid T^0 > t, \boldsymbol{X})$. To assess the covariate effects on the mean residual time, we assume the PMRL model [7] as

$$m(t|\boldsymbol{X}) = m_0(t)\exp(\boldsymbol{\beta}^\top \boldsymbol{X}), \quad (1)$$

where $m_0(t)$ is the unspecified positive baseline mean residual life function and $\boldsymbol{\beta}$ is a $p \times 1$ vector of coefficients. We may use existing methods to fit the model to data from the incident cohort only. [10,11] However, the observed data from the prevalent cohort cannot directly fit model (1) because the survival times are length biased and the backward recurrence times are missing. Under length-biased sampling, it is shown that the conditional density function of the forward recurrence time V^0 given covariates $\boldsymbol{X}$ is

$$f_{V^0|\boldsymbol{X}}(v|\boldsymbol{X}) = \frac{S(v|\boldsymbol{X})}{m(0|\boldsymbol{X})},$$

where $S(\cdot \mid \boldsymbol{X})$ is the conditional survival function of T^0 and $m(0 \mid \boldsymbol{X})$ is the mean survival time of T^0 given $\boldsymbol{X}$. [12] It follows that the hazard function of the forward recurrence time is

$$\lambda^v(t|\boldsymbol{X}) = \frac{S(t|\boldsymbol{X})/m(0|\boldsymbol{X})}{\int_t^\tau S(u|\boldsymbol{X})/m(0|\boldsymbol{X})\mathrm{d}u} = \frac{S(t|\boldsymbol{X})}{\int_t^\tau S(u|\boldsymbol{X})\mathrm{d}u} = \frac{1}{\mathrm{E}(T^0 - t \big| T^0 > t, \boldsymbol{X})} = \frac{1}{m(t|\boldsymbol{X})}$$

where τ is the finite upper bound that satisfies $\Pr(T > \tau) > 0$. Therefore, as discussed by Maguluri and Zhang,[13] Chen and Cheng,[10] and Chen et al.,[11] the PMRL model for T^0 implies the following PH model for the forward recurrence time V^0:

$$\lambda^{v}(t|\boldsymbol{X}) = \{m(t|\boldsymbol{X})\}^{-1} = \{m_0(t)\}^{-1}\exp(-\boldsymbol{\beta}^{\top}\boldsymbol{X}) = \lambda_0^{v}(t)\exp(-\boldsymbol{\beta}^{\top}\boldsymbol{X}), \quad (2)$$

where $\lambda_0^{v}(\cdot)$ is the positive unspecified baseline hazard function of the forward recurrence time.

3 | ESTIMATION METHODS

3.1. | Estimation for Incident Cohort

For data from an incident cohort, Maguluri and Zhang [13] proposed an estimation method under the PMRL model when censoring was absent. Chen et al. [11] extended the method to accommodate right censoring using the inverse probability of censoring weighted (IPCW) approach. The IPCW estimating equation assumes that censoring is independent of the covariates. While the assumption can be relaxed to tackle a censoring distribution that is dependent on the covariates, as discussed in the paper, the censoring mechanism needs to be modelled. An alternative semiparametric estimation procedure was developed based on the counting process theory by Chen and Cheng. [10] We briefly review their method in this section.

Based on the definition of $m(t \mid \boldsymbol{X})$ and using an inversion formula, we can derive the conditional survival function of T^0 given $\boldsymbol{X}$,

$$S(t|\boldsymbol{X}) = \frac{m(0|\boldsymbol{X})}{m(t|\boldsymbol{X})}\exp\left\{-\int_0^t \frac{1}{m(u|\boldsymbol{X})}\mathrm{d}u\right\}.$$

Under model (1), it follows that

$$m_0(t)\mathrm{d}\Lambda_i(t) = \exp(-\boldsymbol{\beta}^{\top}\boldsymbol{X}_i)\mathrm{d}t + \mathrm{d}m_0(t), \quad (3)$$

where $\Lambda_i(t)$ is the cumulative hazard function of T_i^0**.** Let $N_i(t) = I(T_i \le t)\Delta_i$ and $Y_i(t) = I(T_i \ge t)$. Define

$$M_i(t;\boldsymbol{\beta}, m_0) = N_i(t) - \int_0^t Y_i(s)\mathrm{d}\Lambda_i(s;\boldsymbol{\beta}, m_0), \quad (4)$$

where $d\Lambda_i(t;\boldsymbol{\beta}, m_0) = \{\exp(-\boldsymbol{\beta}^\top \boldsymbol{X}_i)dt + dm_0(t)\}$ for $i = 1, \ldots, n_1$.Expression (4) is a zero-mean martingale when $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ and $m_0(\cdot) = m_0^*(\cdot)$, where $\boldsymbol{\beta}^*$ and m_0^* are the true parameter and the true baseline mean function, respectively. Based on equation (3) and expression (4), the following estimating equations are constructed to estimate $m_0(\cdot)$ and $\boldsymbol{\beta}$,

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\left[m_0(t)dN_i(t) - Y_i(t)\{\exp(-\boldsymbol{\beta}^\top \boldsymbol{X}_i)dt + dm_0(t)\}\right] = 0 \quad (5)$$

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\int_0^\tau \boldsymbol{X}_i\left[m_0(t)dN_i(t) - Y_i(t)\{\exp(-\boldsymbol{\beta}^\top \boldsymbol{X}_i)dt + dm_0(t)\}\right] = 0 \quad (6)$$

A closed form solution is available for $m_0(\cdot)$ from equation (5),

$$\widehat{m}_0(t;\boldsymbol{\beta}) = \{\widehat{S}(t)\}^{-1}\int_t^\tau \widehat{S}(u)Q(u;\boldsymbol{\beta})du,$$

where $\widehat{S}(t) = \exp\{-\int_0^t \sum_{i=1}^{n_1} dN_i(u)/\sum_{i=1}^{n_1} Y_i(t)\}$ and $Q(t;\boldsymbol{\beta}) = \sum_{i=1}^{n_1} Y_i(t)\exp(-\boldsymbol{\beta}^\top \boldsymbol{X}_i)/\sum_{i=1}^{n_1} Y_i(t)$. After replacing $m_0(t)$ with $\widehat{m}_0(t;\boldsymbol{\beta})$ in equation (6), we have the estimating function for $\boldsymbol{\beta}$

$$\mathbf{U}_I(\boldsymbol{\beta}) = \frac{1}{n_1}\sum_{i=1}^{n_1}\int_0^\tau \{\boldsymbol{X}_i - \overline{\boldsymbol{X}}(t)\}\{\widehat{m}_0(t;\boldsymbol{\beta})dN_i(t) - Y_i(t)\exp(-\boldsymbol{\beta}^\top \boldsymbol{X}_i)dt\}, \quad (7)$$

where $\overline{\boldsymbol{X}}(t) = \sum_{i=1}^{n_1} Y_i(t)\boldsymbol{X}_i/\sum_{i=1}^{n_1} Y_i(t)$. The estimator $\widehat{\boldsymbol{\beta}}_I$ can be obtained from the solution to $\mathbf{U}_I(\boldsymbol{\beta}) = 0$. Chen and Cheng[10] showed that $n_1^{1/2}(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}^*)$ converges weakly to a normal distribution with mean zero and covariance matrix $A_I^{-1}\sum_I A_I^{-1}$ under the regularity conditions (C1)−(C5) listed in Appendix A.1. We define matrices $\boldsymbol{A}_I$ and Σ_I in Appendix A. 2. The covariance matrix $A_I^{-1}\sum_I A_I^{-1}$ can be consistently estimated by $\{\widehat{A}_I(\widehat{\boldsymbol{\beta}}_I)\}^{-1}\,\widehat{\Sigma}_I\,(\widehat{\boldsymbol{\beta}}_I)\{\widehat{A}_I(\widehat{\boldsymbol{\beta}}_I)\}^{-1}$, where

$$\widehat{\Sigma}_I(\boldsymbol{\beta}) = \frac{1}{n_1}\sum_{i=1}^{n_1}\int_0^{\tau}\{\boldsymbol{X}_i - \bar{\boldsymbol{X}}(t)\}^{\otimes 2} Y_i(t)\widehat{m}_0(t;\boldsymbol{\beta})\{\exp(-\boldsymbol{\beta}^{\top}\boldsymbol{X}_i)\mathrm{d}t + \mathrm{d}\widehat{m}_0(t;\boldsymbol{\beta})\},$$

$$\hat{A}_I(\boldsymbol{\beta}) = \frac{1}{n_1}\sum_{i=1}^{n_1}\int_0^{\tau}\{\boldsymbol{X}_i - \bar{\boldsymbol{X}}(t)\}^{\otimes 2} Y_i(t)\exp(-\boldsymbol{\beta}^{\top}\boldsymbol{X}_i)\mathrm{d}t,$$

in which $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^{\top}$ for any vector $\boldsymbol{a}$.

3.2 | Estimation for Prevalent Cohort

As discussed, data arising from prevalent sampling are subject to length bias, which hinders one from applying the method proposed for the incident cohort. Under the PMRL model, Bai et al. [14] proposed a semiparametric method for right-censored length-biased data, adopting the IPCW approach. That method properly addressed the induced dependent censoring issue and sampling bias, which are commonly encountered in length-biased data with right censoring. However, that method is not directly applicable to our motivating data because the survival times are not available due to missing backward recurrence times. Due to the special relationship between the PMRL and the PH models shown in equation (2), it is sufficient to estimate the covariate effects using only the observed forward recurrence times from the prevalent cohort. Note that we are estimating the same regression coefficient $\boldsymbol{\beta}$ for the target population under model (1) with the prevalent cohort data as with the incident cohort data. This approach has been studied for right-censored length-biased data by Chan et al. [15] for cross-sectional sampled data with no follow-up or data with no information on the disease diagnosis time. The prevalent cohort data in our study belong to the latter case.

Denote $N_i^{v}(t) = I(V_i \leq t)\Delta_i^{v}$ and $Y_i^{v}(t) = I(V_i \geq t)$, for $i = 1, \ldots, n_2$, Define $\boldsymbol{S}^{(k)}(\boldsymbol{\beta}, t) = n^{-1}\sum_{i=1}^{n_2}\boldsymbol{X}_i^{v\otimes k}\exp(-\boldsymbol{\beta}^{\top}\boldsymbol{X}_i^{v})Y_i^{v}(t)$ for k = 0,1, and 2, where $\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}$ for any vector $\boldsymbol{a}$. Based on the relationship shown in equation (2), we can estimate the regression parameter _ by adopting the partial likelihood score function,

$$\mathbf{U}_P(\boldsymbol{\beta}) = \frac{1}{n_2}\sum_{i=1}^{n_2}\int_0^{\tau}\{\boldsymbol{X}_i^{v} - \boldsymbol{\varepsilon}(\boldsymbol{\beta}, t)\}\mathrm{d}N_i^{v}(t), \quad (8)$$

where $\boldsymbol{\varepsilon}(\boldsymbol{\beta}, t) = \boldsymbol{S}^{(1)}(\boldsymbol{\beta}, t)/S^{(0)}(\boldsymbol{\beta}, t)$. The solution to $\mathbf{U}_P(\boldsymbol{\beta}) = 0$ is the estimator $\hat{\boldsymbol{\beta}}_P$ Under the regularity conditions (C1)–(C4), and (C6) listed in Appendix A.1, the distribution of $n_2^{1/2}(\hat{\boldsymbol{\beta}}_P - \boldsymbol{\beta}^{*})$ converges to a normal distribution with mean zero and covariance matrix

$A_P^{-1}\,\Sigma_P\,A_P^{-1}$, where A_P and Σ_P are defined in Appendix A.3. We can consistently estimate $A_P^{-1}\sum_P A_P^{-1}$ by $\{\hat{A}_P(\hat{\boldsymbol{\beta}}_P)\}^{-1}\,\hat{\Sigma}_P\,(\hat{\boldsymbol{\beta}}_P)\{\hat{A}_P(\hat{\boldsymbol{\beta}}_P)\}^{-1}$, where

$$\hat{\Sigma}_P\,(\boldsymbol{\beta}) = \frac{1}{n_2}\sum_{i=1}^{n_2}\left[\int_0^{\tau}\{\boldsymbol{X}_i^{v} - \boldsymbol{\varepsilon}(\boldsymbol{\beta}, t)\}\mathrm{d}N_i^{v}(t)\right]^{\otimes 2},$$

$$\hat{A}_P(\boldsymbol{\beta}) = \frac{1}{n_2}\sum_{i=1}^{n_2}\int_0^{\tau}\left[\frac{\boldsymbol{S}^{(2)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} - \{\boldsymbol{\varepsilon}(\boldsymbol{\beta}, t)\}^{\otimes 2}\right]\mathrm{d}N_i^{v}(t)\,.$$

Note that the estimating function (8) is equivalent to the score function for conventional survival data under the PH model, except for the unknown regression coefficients being negative of $\boldsymbol{\beta}$. Thus, we can implement the estimation method using readily available software.

3.3 | Estimation Using the Combined Cohorts

Although the data arising from the two cohorts have distinct data structures with different time variables, they are from the same target population. Thus, we may use the combined cohort data to make inference for the target cohort under model (1) regarding survival times. To improve statistical efficiency, we propose an estimation method that combines the two weighted estimating functions using data from the incident and prevalent cohorts. We consider a class of weighted linear combinations of the estimating functions (7) and (8):

$$\mathbf{U}_C(\boldsymbol{\beta}) = \frac{1}{n}\{W_1 n_1 \mathbf{U}_I(\boldsymbol{\beta}) + W_2 n_2 \mathbf{U}_p(\boldsymbol{\beta})\} \tag{9}$$

$$= \frac{1}{n}\left[W_1\sum_{i=1}^{n_1}\int_0^{\tau}\{\boldsymbol{X}_i - \bar{\boldsymbol{X}}(t)\}\{\hat{m}_0(t;\boldsymbol{\beta})\mathrm{d}N_i(t) - Y_i(t)\exp(-\boldsymbol{\beta}^{\top}\boldsymbol{X}_i)\mathrm{d}t\}\right.$$

$$\left. + W_2\sum_{i=1}^{n_2}\int_0^{\tau}\{\boldsymbol{X}_i^{v} - \boldsymbol{\varepsilon}(\boldsymbol{\beta}, t)\}\mathrm{d}N_i^{v}(t)\right],$$

where W_1 and W_2 are $p \times p$ weight matrices. We combine the estimating equations derived from each cohort instead of using the weighted average of the two estimators, $\hat{\boldsymbol{\beta}}_I$ and $\hat{\boldsymbol{\beta}}_p$, to avoid imposing a restrictive condition that the optimal estimator is a linear combination of the two estimators. Note that the total sample size increases to $n = n_1 + n_2$ by combining the data from the two cohorts. We can obtain a class of estimators $\tilde{\boldsymbol{\beta}}_C$ by solving $\mathbf{U}_C(\boldsymbol{\beta}) = 0$ for $\boldsymbol{\beta}$.

Among the class of estimators $\tilde{\boldsymbol{\beta}}_C$, we derive the estimator with the smallest asymptotic variance by finding the optimal $\boldsymbol{W} = (W_1, W_2)$. Let $\rho = \lim_{n_1 \to \infty, n_2 \to \infty} n_1/(n_1 + n_2)$. Based on the large sample properties of the estimators $\hat{\boldsymbol{\beta}}_I$ and $\hat{\boldsymbol{\beta}}_p$, the asymptotic covariance matrix of $n^{1/2}\left(\tilde{\boldsymbol{\beta}}_C - \boldsymbol{\beta}*\right)$ is

$$\Omega_C(\boldsymbol{W}) = \{\rho W_1 A_I + (1-\rho) W_2 A_p\}^{-1} \{\rho W_1 \sum_I W_1^T + (1-\rho) W_2 \sum_p W_2^T\} \left[\{\rho W_1 A_I + (1-\rho) W_2 A_p\}^{-1}\right]^{\mathrm{T}}.$$

By the matrix Cauchy–Schwarz inequality,[16] for any $\boldsymbol{W}$,

$$\Omega_C(\boldsymbol{W}) \geq \Omega_{opt} = \{\rho A_I \sum_I^{-1} A_I + (1-\rho) A_P \sum_P^{-1} A_P\}^{-1}.$$

We can attain the efficiency bound Ω_{opt} when the weight matrices $W_1 = A_I \sum_I^{-1}$ and $W_2 = A_p \sum_p^{-1}$, which are the optimal weights. Since the optimal weights depend on the unknown parameter $\boldsymbol{\beta}$, we proceed to a two-step estimation. We first derive an estimator that is consistent with $\boldsymbol{\beta*}$ by solving $\mathbf{U}_C(\boldsymbol{\beta}) = 0$ with $W_1 = W_2 = \boldsymbol{I}_{p \times p}$, where $\boldsymbol{I}_{p \times p}$ is the identity matrix, to obtain the first-step estimator $\hat{\boldsymbol{\beta}}_C$. Then, the efficient estimator $\hat{\boldsymbol{\beta}}_{opt}$ is the solution to

$$\mathbf{U}_{opt}(\boldsymbol{\beta}) = \frac{1}{n}\{\widehat{W}_1 n_1 U_I(\boldsymbol{\beta}) + \widehat{W}_2 n_2 U_p(\boldsymbol{\beta})\} = 0,$$

where $\widehat{W}_1 = \hat{A}_I(\hat{\boldsymbol{\beta}}_C)\{\hat{\Sigma}_I(\hat{\boldsymbol{\beta}}_C)\}^{-1}$ and $\widehat{W}_2 = \hat{A}_P(\hat{\boldsymbol{\beta}}_C)\{\hat{\Sigma}_P(\hat{\boldsymbol{\beta}}_C)\}^{-1}$. The asymptotic properties of $\hat{\boldsymbol{\beta}}_{oPt}$ are summarized in the following theorem.

Theorem 1. Under the regularity conditions listed in Appendix A.1, $n^{1/2}(\hat{\boldsymbol{\beta}}_{opt} - \boldsymbol{\beta}^*)$ converges weakly to a normal distribution with mean zero and covariance matrix Ω_{opt}.

The detailed proofs of Theorem 1 are provided in Appendix A.4. The covariance matrix Ω_{opt} can be consistently estimated $\widehat{\Omega}_{opt}$,

$$\left[\hat{\rho}\hat{A}_I(\hat{\boldsymbol{\beta}}_{opt})\{\hat{\Sigma}_I(\hat{\boldsymbol{\beta}}_{opt})\}^{-1}\hat{A}_I(\hat{\boldsymbol{\beta}}_{opt}) + (1-\hat{\rho})\hat{A}_P(\hat{\boldsymbol{\beta}}_{opt})\{\hat{\Sigma}_P(\hat{\boldsymbol{\beta}}_{opt})\}^{-1}\hat{A}_P(\hat{\boldsymbol{\beta}}_{opt})\right]^{-1},$$

where $\hat{\rho} = n_1/n$.

4 | SIMULATION STUDY

We conducted simulation studies to investigate the finite sample properties of the proposed estimation method for the combined cohort data. We simulated 1000 datasets that consist of

n_1 subjects from the incident cohort and n_2 subjects from the prevalent cohort. Total sample sizes of $n = n_1 + n_2 = 200$ and 400 were considered with various combinations. We considered two covariates: X_1 from a Bernoulli distribution with probability 0.5 and X_2 from a uniform distribution (0, 1) for both cohorts. Conditioning on X_1 and X_2, the survival time T^0 was generated from the same target population under the mean residual life model $m(t \mid X_1, X_2) = (at+b)\exp(\beta_1 X_1 + \beta_2 X_2)$, where parameters for the baseline mean function $(a, b) = (0.1, 0.5)$ and the true coefficients $(\beta_1, \beta_2) = (0.5, -0.5)$. For the incident cohort, we randomly generated n_1 observations, (T_i^0, X_{1i}, X_{2i}), $i = 1, \dots, n_1$. For the prevalent cohort, we generated the left truncation time A from a uniform distribution and only kept observations that satisfy $T^0 > A$. We continued the sampling procedure until we sampled n_2 observations $(V_j^0, X_{1j}^v, X_{2j}^v)$ $j = 1, \dots, n_2$, where $V_j^0 = T_j^0 - A_j$, and $X_{1j}^v = X_{1j}, X_{2j}^v = X_{2j}$ for subject j with $T_j^0 > A_j$. Since both cohorts are subject to right censoring, we generated censoring times C and C^v from a uniform distribution $(0, \tau_C)$ and chose τ_C to allow for 15% and 30% of censoring rates overall. Under this setting, the censoring rate of each cohort is about the same. The distributions of C and C^v share the same support because the follow-up periods for both cohorts are the same in practice. The generated dataset consists of $\{(T_i, \Delta_i, X_{1i}, X_{2i}), (V_j, \Delta_j^v, X_{1j}^v, X_{2j}^v); i = 1, \dots, n_1, j = 1, \dots, n_2\}$.

We denoted $\hat{\boldsymbol{\beta}}_I$ as the estimator using the simulated incident cohort data only, $\hat{\boldsymbol{\beta}}_P$ using the simulated prevalent cohort data only, and $\hat{\boldsymbol{\beta}}_C$ and $\hat{\boldsymbol{\beta}}_{opt}$ as the proposed estimators using data from both cohorts with identity weight matrices and the optimal weights, respectively. Tables 1 and 2 summarize the simulation results. When the overall censoring rate is as low as 15%, all estimators present virtually unbiased point estimates, the asymptotic standard errors are close to the empirical standard deviations of the point estimates, and the coverage probabilities are close to the nominal level of 95%. We note that the relative efficiency of the estimators $\hat{\boldsymbol{\beta}}_I$ and $\hat{\boldsymbol{\beta}}_P$ highly depends on the number of samples in each cohort. When there are more samples and hence more failure events in the incident cohort than in the prevalent cohort (i.e., $n_1 > n_2$), $\hat{\boldsymbol{\beta}}_I$ has smaller variance, which indicates that it is more efficient than $\hat{\boldsymbol{\beta}}_P$, and vice versa. When the proposed method is used for the combined cohort data, we have an increased sample size of $n_1 + n_2$. Thus, we observe smaller variance estimates for $\hat{\boldsymbol{\beta}}_C$ and $\hat{\boldsymbol{\beta}}_{opt}$ compared to $\hat{\boldsymbol{\beta}}_I$ and $\hat{\boldsymbol{\beta}}_P$ under all settings. To assess the efficiency gain of the proposed estimators over $\hat{\boldsymbol{\beta}}_I$ and $\hat{\boldsymbol{\beta}}_P$, we compute the relative efficiency, which is defined as the ratio of the mean squared errors of the estimators. For example, when $n_1 = 100$, $n_2 = 100$, and the censoring rate is 15%, $\hat{\boldsymbol{\beta}}_C$ for β_1 is 1.86 and 1.93 times more efficient than $\hat{\boldsymbol{\beta}}_I$ and $\hat{\boldsymbol{\beta}}_P$, respectively; and $\hat{\boldsymbol{\beta}}_{opt}$ is respectively 2.08 and 2.15 times more efficient. The proposed estimator with optimal weights $\hat{\boldsymbol{\beta}}_{opt}$ is relatively more efficient than $\hat{\boldsymbol{\beta}}_C$ across all settings. While the point estimates for $\hat{\boldsymbol{\beta}}_{opt}$ tend to be slightly more biased than $\hat{\boldsymbol{\beta}}_C$ due to the two-step estimation procedure, the mean squared errors of $\hat{\boldsymbol{\beta}}_{opt}$ are smaller in every setting.

With an increased censoring rate of 30%, we find some bias for $\hat{\boldsymbol{\beta}}_I$, where only the incident cohort data are used. A similar trend was observed in the original simulation studies on $\hat{\boldsymbol{\beta}}_I$ conducted by Chen and Cheng. [10] In the simulation results under a censoring rate of 30%, we observe that the estimators $\hat{\boldsymbol{\beta}}_C$ and $\hat{\boldsymbol{\beta}}_{opt}$ are less biased and more efficient than $\hat{\boldsymbol{\beta}}_I$. Therefore, combining information from the prevalent cohort data with that from the incident cohort data is desirable, especially under heavy censoring rates.

5 | APPLICATION

The Nun Study, introduced in Section 1, has been conducted to examine risk factors for the progression of dementia, with a cohort of 678 members of the School Sisters of Notre Dame religious congregation who were 75 years of age or older and recruited between 1991 and 1993. [1] Each participant received an assessment of her cognitive and physical function near-annually up to 10 years. At each examination, the participant's cognitive status was recorded as one of the five following states: cognitively intact for age, cognitive deficit that does not affect activities of daily living, cognitive deficit in one or more activities of daily living, clinical dementia, and death. Covariates such as age at each exam, presence of the apolipoprotein E-ϵ4 allele (APOE4), and the level of education were collected.

To illustrate the proposed estimation method, we use the combined cohort data, which consist of 501 subjects with complete data from the Nun Study. Among them, 153 incident and 77 prevalent cases were used in the analysis. In the data, the exact time of death was recorded if it occurred before the last follow-up. If a subject did not die by the last examination, her survival time was censored. Among the incident cases, 29 (19%) subjects were right censored; and only two (2.6%) were right censored among the prevalent cases. The overall censoring rate was as low as 13.5%. For the incident cohort, the data include the survival time from dementia diagnosis until death or the censoring event. When a subject was assessed as clinically demented at one of the annual examinations, we assumed that dementia occurred in the middle of two consecutive examinations. However, for the prevalent cohort data, we only have the information that the subject was demented prior to enrollment; hence, the backward recurrence time is missing. Instead, we have the forward recurrence times from study enrollment until death or the censoring event for the prevalent cohort. We considered two covariates of interest: the level of education and the presence of the genetic risk factor APOE4. The distribution of the covariates are summarized in Table 3 by each cohort and for the combined cohorts.

We conducted regression analyses to estimate the effects of the educational level and APOE4 on the mean residual survival time under the PMRL model (1). The analyses were carried out using the incident cohort only, the prevalent cohort only, and the combined cohort data with optimal weights. In the analysis of the incident cohort data, the support of the censoring distribution is greater than that of the survival distribution, which satisfies the assumption for the method using only the incident cohort. [10] The estimated distributions of the survival time and the censoring time are provided as Figure S2 in the web-based supplementary materials. We present the results in Table 4 . None of the estimated regression parameters were found to be significantly associated with the mean residual survival time, which is consistent with the findings in the literature. Qiu et al. [17] and Helmer et al. [18] showed that educational level was not significantly correlated with the mortality of subjects who had dementia, while a lower level of education was found to be associated with higher risk of dementia in other studies. [19] Mez et al. [20] suggested that the incidence of dementia may mediate the effect of APOE4 on mortality, given that both APOE4 and dementia are high risk factors for decreased survival times among older adults. Thus, among subjects diagnosed with dementia,

APOE4 has not been found to be a significant risk factor for death.

Under the assumption that the incident and prevalent cohorts are from the same population, we can examine the proportional means assumption by checking the proportional hazards assumption using the prevalent cohort. We confirmed that the assumption is reasonable: the *p*-values are 0.66 and 0.19 for the presence of APOE4 and the level of education, respectively; and 0.39 for the global test. However, it should be noted that the model diagnostic test may have low power. Another assumption is that data observed in the prevalent cohort are subject to length bias (i.e., the incidence of dementia follows a stationary Poisson process). However, the information about the time from dementia onset to enrollment is missing for the prevalent cohort data, which hinders one from checking this assumption. As an alternative, we can compute the dementia incidence rate using the incident cohort only data, provided that the two cohorts are from the same population. The incidence rate was fairly constant over the follow-up period, with no specific trend. Hence, the stationarity assumption is reasonable for our application.

6 | CONCLUSION

In observational studies, prevalent samples are commonly collected along with the incident cohort from a single study population. Combining data from incident and prevalent cohorts can substantially improve efficiency and ensure robustness of the estimators when assessing the risk factors' effects on survival times. This is an efficient way of utilizing the data because the combined cohort data are usually available at no additional cost. While statistical methods for the analysis of the combined cohort data would make invaluable contributions to many studies, such methods are limited in the literature.

In this paper, we assume the PMRL model for the target population. One advantage of assuming such a model is that it directly leads to the PH model on the forward recurrence times for the prevalent cohort. Hence, we can use the conventional survival method for the incident cohort under the PMRL model. For the prevalent cohort data, which has a nonstandard structure with missing backward recurrence times, we can use the PH model without additional assumptions or extra effort. Thus, the proposed estimation method involves two estimating functions that are constructed differently for data from the incident and prevalent cohorts.

In the estimating function for the combined data (9), we only use data from the incident cohort to derive the consistent estimator $\widehat{m}_0(t;\boldsymbol{\beta})$ for $m_0(t)$. To estimate the baseline mean function $m_0(t)$ more efficiently, one may consider combining the data om the prevalent cohort. Based on equation (2), $m_0(t)$ is the inverse of the baseline hazard function of the forward recurrence time,$\lambda_0^{v}(t)$. Hence, a naive approach is to estimate the inverse of $\lambda_0^{v}(t)$ based on the Nelson–Aalen estimator for the cumulative hazard ^{0}function. However, the estimated baseline cumulative hazar^{0}d function is nonsmooth and results in a noisy estimator for $\lambda_0^{v}(t)$ as in conventional survival analyses, which leads to an unstable estimation of $m_0(t)$. As an alternative, one may adopt the kernel smoothing method to estimate the baseline hazard function,$\lambda_0^{v}(t)$.[21] A major drawback of applying the smoothing method is that the choice of bandwidth, which is crucial, involves computationally intensive procedures. Further studies on combining data for more efficient estimation of $m_0(t)$ are of interest.

Subjects were examined periodically in the Nun Study. While the dates of death were accurately recorded, the onset of dementia is only known to occur within time intervals (i.e., interval censored). In our application, we adopted a simple approach by assuming that the event occurred in the middle of the interval since that was not the focus of the current paper. Further research that tackles the issue of interval censoring is certainly warranted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The work was partially supported by the U.S. National Institutes of Health through grants CA193878 and CA016672. The Nun Study data reported in this article were collected from the SMART project (AG386561). The authors also acknowledge the Texas Advanced Computing Center at The University of Texas at Austin for providing HPC resources that contributed to the research results reported within this paper.

APPENDIX

A LARGE SAMPLE PROPERTIES OF THE ESTIMATORS

A.1 Regularity conditions

(C1) Given any $\boldsymbol{X} = \boldsymbol{x}, \Pr(T^{0} < C/\mathrm{x}) > 0$; and given any $\boldsymbol{X}^{v} = \boldsymbol{x}, \Pr(V^{0} < C^{v}/\boldsymbol{x}) > 0$.

(C2) The parameter space of $\boldsymbol{\beta}$ is a compact subset of $\mathbb{R}^{p}$, and the true parameter value $\boldsymbol{\beta}^{*}$ is in the interior of the parameter space.

(C3) The true baseline mean function $m_0^{*}(t)$ is continuously differentiable on $[0, \boldsymbol{\tau}]$.

(C4) A $p \times 1$ vector of covariates $\boldsymbol{X}$ is bounded by some constant, and not contained in a $(p - 1)$-dimensional hyperplane.

(C5) $A_I = \int_0^{\tau} \mathrm{E}\left[\{\boldsymbol{X} - \mu_X(t)\}^{\otimes 2} S^{*}(t/\boldsymbol{X})\exp(-\boldsymbol{\beta}^{*\mathrm{T}}\boldsymbol{X})\right] \mathrm{d}t$ is nonsingular, where $\mu_X(t)$ is the limit of $\overline{\boldsymbol{X}}(t)$ as $n_1 \to \infty$ and $S^{*}(t/\boldsymbol{X}) = \Pr(T > t/\boldsymbol{X})$.

(C6) $A_P = \int_0^\tau \left[\frac{\boldsymbol{s}^{(2)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)} - \{\boldsymbol{e}(\beta^*, t)\}^{\otimes 2} \right] \boldsymbol{s}^{(0)}(\boldsymbol{\beta}^*, t)\lambda_0^{v}(t)\, \mathrm{d}\boldsymbol{t}$ is positive definite, where $\boldsymbol{s}^{(k)}(\boldsymbol{\beta}, t)$ is the limit of $\boldsymbol{S}^{(k)}(\boldsymbol{\beta}, t)$ for $k = 0,1$, and 2, and $\boldsymbol{e}(\boldsymbol{\beta}, t)$ is the limit of $\boldsymbol{\varepsilon}(\boldsymbol{\beta}, t)$ as $n_2 \to \infty$.

A.2 Asymptotic properties of $\hat{\boldsymbol{\beta}}_I$

The asymptotic properties of the estimator $\hat{\boldsymbol{\beta}}_I$ have been established in the appendix of Chen and Cheng. 10 Here, we briefly outline the results. Given that $\widehat{m}_0(t; \boldsymbol{\beta}^*)$ converges to $m_0^*(t)$ almost surely, we have

$$n_1^{1/2}\mathbf{U}_I(\boldsymbol{\beta}^*) = n_1^{-1/2}\sum_{i=1}^{n_1}\int_0^\tau (\boldsymbol{X}_i - \bar{\boldsymbol{X}}(t) - \frac{\mathrm{E}\left[S(t/\boldsymbol{X})\{\boldsymbol{X} - \mu_X(t)\}\right]}{\mathrm{E}\{S(t/\boldsymbol{X})\}})\, m_0^*(t)\mathrm{d}M_i(t) + o_P(1)\,.$$

Thus $n_1^{1/2}\mathbf{U}_I(\boldsymbol{\beta}^*)$ converges weakly to a normal distribution with mean zero and covariance matrix

$$\sum\nolimits_I = \int_0^\tau \mathrm{E}\left[\{\boldsymbol{X} - \mu_X(t)\}^{\otimes 2} S^*(t/\boldsymbol{X})m_0^*(t)\{\exp(-\boldsymbol{\beta}^{*\mathrm{T}}\boldsymbol{X})\mathrm{d}t + \mathrm{d}m_0^*(t)\}\right].$$

Provided that $\partial\widehat{m}_0(t, \boldsymbol{\beta}^*)/\partial\beta = -m_0^*(t)\mu_X(t) + o_P(1)$, it is shown that $\partial U_I(\boldsymbol{\beta}^*)/\partial\boldsymbol{\beta}$ converges in probability to $\boldsymbol{A}_I$, is defined in (C5). By applying the Taylor series expansion, one can show that $n_1^{1/2}U_I(\boldsymbol{\beta}^*) = \{-\partial U_I(\boldsymbol{\beta}^*)/\partial\boldsymbol{\beta}\}n_1^{1/2}(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}) + o_P(1)$. Hence, it follows that $n_1^{1/2}(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta})$ converges weakly to a normal distribution with mean zero and covariance matrix $A_I^{-1}\sum_I A_I^{-1}$.

A.3 Asymptotic properties of $\hat{\boldsymbol{\beta}}_P$

We establish the asymptotic properties of $\hat{\boldsymbol{\beta}}_P$ following the large sample studies conducted by Andersen and Gill [22] for conventional survival data. Let $M_i^{v}(t) = N_i^{v}(t) - \int_0^t \lambda_i^{v}(s)\,\mathrm{d}s$, where $\lambda_i^{v}(t) = \lambda_0^{v}(t)\exp(-\boldsymbol{\beta}^\top\boldsymbol{X}_i)$. We can represent the estimating function $\mathbf{U}_P(\boldsymbol{\beta})$ evaluated at $\boldsymbol{\beta}^*$ as follow:

$$n_2^{1/2}\mathbf{U}_P(\boldsymbol{\beta}^*) = n_2^{-1/2}\sum_{i=1}^{n_2}\int_0^\tau \{\boldsymbol{X}_i^{v} - \varepsilon(\boldsymbol{\beta}^*, t)\}\mathrm{d}M_i^{v}(t) + o_P(1)\,.$$

The distribution of $n_2^{1/2}\mathbf{U}_P(\boldsymbol{\beta}^*)$ is asymptotically normal with mean zero and covariance matrix

$$\sum_P = \int_0^\tau \left[\frac{\mathrm{s}^{(2)}(\boldsymbol{\beta}^*, t)}{\boldsymbol{s}^{(2)}(\boldsymbol{\beta}^*, t)} - \{\boldsymbol{e}(\boldsymbol{\beta}^*, t)\}^{\otimes 2} \right] \boldsymbol{s}^{(0)}(\boldsymbol{\beta}^*, t)\lambda_0^{\upsilon}(t)\mathrm{d}t \,.$$

By the Taylor series expansion,$n_1^{1/2}\mathbf{U}_P(\boldsymbol{\beta}^*) = \{-\partial \mathbf{U}_P(\boldsymbol{\beta}^*)/\partial \boldsymbol{\beta}\} n_2^{1/2}(\hat{\boldsymbol{\beta}}_P - \boldsymbol{\beta}^*) + o_P(1)$. Note that $\partial U_P(\boldsymbol{\beta}^*)/\partial \boldsymbol{\beta}$ converges in probability to A_P, which is defined in (C6). Thus, $n_2^{1/2}(\hat{\boldsymbol{\beta}}_P - \boldsymbol{\beta}^*)$ asymptotically follows a normal distribution with mean zero and covariance matrix $A_P^{-1} \sum_P A_P^{-1}$.

A.4 Proofs of Theorem 3.1

Given the optimal weights $W_1 = A_I\, \Sigma_I^{-1}$ and $W_2 = A_P\, \Sigma_P^{-1}$, we rewrite the estimating function $\mathbf{U}_{opt}(\boldsymbol{\beta})$ in summations of i.i.d. vectors, as follows.

$$n^{1/2}\mathbf{U}_{opt}(\boldsymbol{\beta}) = \sqrt{\frac{n_1}{n}} A_I \sum\nolimits_I^{-1} \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \int_0^\tau (\boldsymbol{X}_i - \bar{\boldsymbol{X}}(t) - \frac{\mathrm{E}\left[S(t/\boldsymbol{X})\{\boldsymbol{X} - \mu_X(t)\}\right]}{\mathrm{E}\{S(t/\boldsymbol{X})\}}) \quad m_0^*(t)\mathrm{d}M_i(t)$$
$$+ \sqrt{\frac{n_2}{n}} A_P \sum\nolimits_P^{-1} \frac{1}{\sqrt{n_2}} \sum_{i=1}^{n_2} \int_0^\tau \{\boldsymbol{X}_i^{\upsilon} - \boldsymbol{\varepsilon}(\boldsymbol{\beta}, t)\}\mathrm{d}M_i^{\upsilon}(t) + o_P(1)\,.$$

Based on the asymptotic properties of $\hat{\boldsymbol{\beta}}_I$ and $\hat{\boldsymbol{\beta}}_P$, it follows that $n^{1/2}\mathbf{U}_{opt}(\boldsymbol{\beta}^*)$ is asymptotically normal with mean zero and covariance matrix $\Sigma \;= \rho A_I\, \Sigma_I^{-1}\; A_I + (1-\rho)A_P\, \Sigma_P^{-1}\; A_P$. This is straightforward because the incident and prevalent cohorts are independent.

By the Taylor series expansion of $\mathbf{U}(\hat{\boldsymbol{\beta}}_{opt})$ around $\boldsymbol{\beta}$*, we have

$$n^{1/2}(\hat{\boldsymbol{\beta}}_{opt} - \boldsymbol{\beta}^*) = \{-\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{U}_{opt}(\bar{\boldsymbol{\beta}})\}^{-1} n^{1/2}\mathbf{U}_{opt}(\boldsymbol{\beta}^*),$$

where $\bar{\boldsymbol{\beta}}$ is on the line segment between $\hat{\boldsymbol{\beta}}_{\mathrm{opt}}$ and $\boldsymbol{\beta}$*, and

$$\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{U}_{opt}(\bar{\boldsymbol{\beta}}) = \frac{n_1}{n} A_I \sum\nolimits_I^{-1} \frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{U}_I(\bar{\beta}) + \frac{n_2}{n} A_P \sum\nolimits_P^{-1} \frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{U}_P(\bar{\beta})\,.$$

We can easily show that $\partial \mathbf{U}_{opt}(\bar{\boldsymbol{\beta}})/\partial \boldsymbol{\beta}$ converges in probability to $\rho A_I\, \Sigma_I^{-1}\; A_I + (1-\rho)A_P\, \Sigma_P^{-1}\; A_P$, which is equal to Σ. Therefore, $n^{1/2}(\hat{\boldsymbol{\beta}}_{opt} - \boldsymbol{\beta}^*)$ is asymptotically normal with mean zero and covariance $\Omega_{opt} = \;\Sigma^{-1}\;\; \Sigma\;\; \Sigma^{-1}\;\; = \;\Sigma$.

Denote an arbitrarily small neighborhood of $\boldsymbol{\beta}^*$ as $\mathcal{B}$ Following the arguments in Chen and Cheng,[10] $\Pr(\hat{\boldsymbol{\beta}}_{opt} \in B) = 1$ because $\mathbf{U}_{opt}(\boldsymbol{\beta}^*) \rightarrow 0$ can be extended to any $\boldsymbol{\beta} \in \mathcal{B}$ under the regularity conditions on uniform convergence. Thus, $\hat{\boldsymbol{\beta}}_{opt}$ is consistent with $\boldsymbol{\beta}^*$**.**

Given the consistency of $\hat{A}_I(\boldsymbol{\beta})$, $\hat{A}_P(\boldsymbol{\beta})$, $\widehat{\Sigma}_I(\boldsymbol{\beta})$, and $\widehat{\Sigma}_P(\boldsymbol{\beta})$, and assuming that Σ_I and Σ_P are nonsingular, we can show that the estimators of the optimal weights $\widehat{W}_1 = \hat{A}_I(\hat{\beta}_C)\{\widehat{\Sigma}_I(\hat{\beta}_C)\}^{-1}$ and $\widehat{W}_2 = \hat{A}_P(\hat{\beta}_C)\{\widehat{\Sigma}_P(\hat{\beta}_C)\}^{-1}$ converge in probability to $A_I \Sigma_I^{-1}$ and $A_P \Sigma_P^{-1}$, respectively, where $\hat{\boldsymbol{\beta}}_C$ is a consistent estimator of $\boldsymbol{\beta}^*$

References

1. Snowdon DA, Greiner LH, Mortimer JA, Riley KP, Greiner PA, Markesbery WR. Brain infarction and the clinical expression of Alzheimer disease. The Nun Study. JAMA 1997;277:813–817. [PubMed: 9052711]
2. Tyas SL, Salazar JC, Snowdon DA, et al. Transitions to mild cognitive impairments, dementia, and death: findings from the Nun Study. Am J Epidemiol 2007;165:1231–1238. [PubMed: 17431012]
3. Yu L, Tyas SL, Snowdon DA, Kryscio RJ. Effects of ignoring baseline on modeling transitions from intact cognition to dementia. Comput Stat Data Anal 2009;53:3334–3343. [PubMed: 20161282]
4. Yu L, Griffith WS, Tyas SL, Snowdon DA, Kryscio RJ. A nonstationary Markov transition model for computing the relative risk of dementia before death. Stat Med 2010;.
5. Wei S, Xu L, Kryscio RJ. Markov transition model to dementia with death as a competing event. Comput Stat Data Anal 2014;80:78–88. [PubMed: 25110380]
6. Wei S, Kryscio RJ. Semi-Markov models for interval censored transient cognitive states with back transitions and a competing risk. Stat Methods Med Res 2016;25:2909–2924. [PubMed: 24821001]
7. Oakes D, Dasu T. A note on residual life. Biometrika 1990;77:409–410.
8. Addona V, Wolfson DB. A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. Lifetime Data Anal 2006;12:267–284. [PubMed: 16917734]
9. Asgharian M, Wolfson DB, Zhang X. Checking stationarity of the incidence rate using prevalent cohort survival data. Stat Med 2006;25:1751–1767. [PubMed: 16220462]
10. Chen YQ, Cheng S. Semiparametric regression analysis of mean residual life with censored survival data. Biometrika 2005;92:19–29.
11. Chen YQ, Jewell NP, Lei X, Cheng SC. Semiparametric estimation of proportional mean residual life model in presence of censoring. Biometrics 2005;61:170–178. [PubMed: 15737090]
12. Cox DR. Renewal Theory London: Methuen; 1962.
13. Maguluri G, Zhang CH. Estimation in the mean residual life regression model. J R Stat Soc Series B Stat Method 1994;56:477–489.
14. Bai F, Huang J, Zhou Y. Semiparametric inference for the proportional mean residual life model with right-censored length-biased data. Stat Sin 2016;26:1129–1158.
15. Chan KCG, Chen YQ, Di CZ. Proportional mean residual life model for right-censored length-biased data. Biometrika 2012;99:995–1000. [PubMed: 23843676]
16. Chaganty NR, Joe H. Efficiency of generalized estimating equations for binary responses.. J R Stat Soc Series B Stat Method 2004;66:851–860.
17. Qiu C, Bäckman L, Winblad B, Agüero-Torres H, Fratiglioni L. The influence of education on clinically diagnosed dementia incidence and mortality data from the Kungsholmen Project. Arch Neurol 2001;58:2034–2039. [PubMed: 11735777]
18. Helmer C, Joly P, Letenneur L, Commenges D, Dartigues JF. Mortality with dementia: results from a French prospective community-based cohort. Am J Epidemiol 2001;154:642–648. [PubMed: 11581098]

19. Sharp ES, Gatz M. The relationship between education and dementia: an updated systematic review. Alzheimer Dis Assoc Disord 2011;25:289–304. [PubMed: 21750453]

20. Mez J, Marden JR, Mukherjee S, et al. Alzheimer's disease genetic risk variants beyond APOEe4 predict mortality. Alzheimers Dement 2017;doi: 10.1016/j.dadm.2017.07.002.

21. Wells MT. Nonparametric kernel estimation in counting processes with explanatory variables. Biometrika 1994;81:759– 801.

22. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. Ann Stat 1982;10:1100–1120.

TABLE 1

Summary statistics of simulation results for estimating (β_1, β_2) = (0.5, −0.5) with n = 200. Monte Carlo mean of the estimates (Est), the empirical standard deviation (SD), the mean standard error (SE), the mean squared error (MSE) and the coverage probability (CP) using incident cohort only ($\hat{\beta}_I$), prevalent cohort only ($\hat{\beta}_P$), and both incident and prevalent cohorts ($\hat{\beta}_C$ and $\hat{\beta}_{opt}$) with sample sizes of n_1 and n_2 for incident and prevalent cohorts, respectively, and censoring rates (cr) of 15% and 30%.

				β_1					B_2				
n_1	n_2	cr		Est	SD	SE	MSE	CP	Est	SD	SE	MSE	CP
125	75	15%	$\hat{\beta}_I$	0.481	0.208	0.206	0.043	0.943	−0.489	0.347	0.359	0.120	0.950
			$\hat{\beta}_P$	0.526	0.276	0.274	0.077	0.948	−0.502	0.492	0.468	0.242	0.950
			$\hat{\beta}_C$	0.502	0.171	0.172	0.029	0.954	−0.484	0.302	0.297	0.091	0.952
			$\hat{\beta}_{opt}$	0.487	0.162	0.164	0.026	0.950	−0.493	0.279	0.282	0.078	0.947
		30%	$\hat{\beta}_I$	0.422	0.199	0.198	0.046	0.934	−0.429	0.336	0.346	0.118	0.936
			$\hat{\beta}_P$	0.519	0.306	0.300	0.094	0.951	−0.499	0.547	0.517	0.299	0.940
			$\hat{\beta}_C$	0.471	0.186	0.181	0.035	0.948	−0.450	0.322	0.314	0.106	0.934
			$\hat{\beta}_{opt}$	0.443	0.164	0.165	0.030	0.940	−0.456	0.281	0.285	0.081	0.934
100	100	15%	$\hat{\beta}_I$	0.476	0.232	0.229	0.054	0.945	−0.470	0.390	0.401	0.153	0.950
			$\hat{\beta}_P$	0.513	0.236	0.235	0.056	0.951	−0.524	0.401	0.399	0.161	0.938
			$\hat{\beta}_C$	0.498	0.169	0.171	0.029	0.955	−0.499	0.286	0.294	0.081	0.951
			$\hat{\beta}_{opt}$	0.486	0.161	0.163	0.026	0.950	−0.497	0.273	0.280	0.074	0.944
		30%	$\hat{\beta}_I$	0.417	0.222	0.222	0.056	0.927	−0.416	0.381	0.387	0.152	0.938
			$\hat{\beta}_P$	0.509	0.256	0.258	0.066	0.951	−0.539	0.446	0.442	0.200	0.952
			$\hat{\beta}_C$	0.476	0.182	0.182	0.034	0.942	−0.486	0.307	0.314	0.095	0.958
			$\hat{\beta}_{opt}$	0.450	0.165	0.167	0.030	0.942	−0.473	0.284	0.288	0.081	0.953
75	125	15%	$\hat{\beta}_I$	0.473	0.263	0.263	0.070	0.943	−0.469	0.463	0.459	0.215	0.933
			$\hat{\beta}_P$	0.512	0.217	0.209	0.047	0.939	−0.514	0.365	0.353	0.134	0.950
			$\hat{\beta}_C$	0.501	0.172	0.170	0.030	0.940	−0.495	0.297	0.289	0.088	0.947
			$\hat{\beta}_{opt}$	0.489	0.163	0.163	0.027	0.946	−0.492	0.286	0.277	0.082	0.943
		30%	$\hat{\beta}_I$	0.421	0.252	0.254	0.069	0.942	−0.415	0.452	0.443	0.212	0.940

				β_1					B_2				
n_1	n_2	cr		**Est**	**SD**	**SE**	**MSE**	**CP**	**Est**	**SD**	**SE**	**MSE**	**CP**
			$\hat{\beta}_P$	0.515	0.234	0.227	0.055	0.946	−0.512	0.407	0.388	0.166	0.945
			$\hat{\beta}_C$	0.491	0.185	0.181	0.034	0.944	−0.478	0.326	0.311	0.107	0.953
			$\hat{\beta}_{opt}$	0.466	0.167	0.168	0.029	0.949	−0.469	0.307	0.289	0.095	0.944

TABLE 2

Summary statistics of simulation results for estimating (β_1, β_2) = (0.5, −0.5) with n = 400. Monte Carlo mean of the estimates (Est), the empirical standard deviation (SD), the mean standard error (SE), the mean squared error (MSE) and the coverage probability (CP) using incident cohort only ($\hat{\beta}_I$), prevalent cohort only ($\hat{\beta}_P$), and both incident and prevalent cohorts ($\hat{\beta}_C$ and $\hat{\beta}_{opt}$) with sample sizes of n_1 and n_2 for incident and prevalent cohorts, respectively, and censoring rates (cr) of 15% and 30%.

				β_1					B_2				
n_1	n_2	cr		Est	SD	SE	MSE	CP	Est	SD	SE	MSE	CP
250	150	15%	$\hat{\beta}_I$	0.479	0.142	0.146	0.021	0.946	−0.472	0.252	0.254	0.064	0.957
			$\hat{\beta}_P$	0.514	0.184	0.190	0.034	0.970	−0.529	0.334	0.322	0.113	0.949
			$\hat{\beta}_C$	0.496	0.116	0.120	0.013	0.954	−0.496	0.206	0.207	0.042	0.954
			$\hat{\beta}_{opt}$	0.488	0.111	0.115	0.012	0.955	−0.493	0.196	0.199	0.038	0.960
		30%	$\hat{\beta}_I$	0.424	0.138	0.141	0.025	0.916	−0.422	0.243	0.246	0.065	0.941
			$\hat{\beta}_P$	0.509	0.204	0.208	0.042	0.962	−0.534	0.371	0.357	0.138	0.941
			$\hat{\beta}_C$	0.467	0.124	0.126	0.017	0.947	−0.473	0.217	0.219	0.048	0.957
			$\hat{\beta}_{opt}$	0.446	0.113	0.116	0.016	0.927	−0.460	0.198	0.201	0.041	0.953
200	200	15%	$\hat{\beta}_I$	0.475	0.162	0.163	0.027	0.945	−0.469	0.280	0.284	0.079	0.947
			$\hat{\beta}_P$	0.516	0.172	0.164	0.030	0.931	−0.508	0.272	0.278	0.074	0.955
			$\hat{\beta}_C$	0.501	0.121	0.120	0.015	0.940	−0.490	0.199	0.206	0.040	0.961
			$\hat{\beta}_{opt}$	0.490	0.115	0.115	0.013	0.951	−0.488	0.191	0.198	0.037	0.956
		30%	$\hat{\beta}_I$	0.420	0.157	0.157	0.031	0.914	−0.420	0.273	0.275	0.081	0.941
			$\hat{\beta}_P$	0.516	0.188	0.179	0.035	0.938	−0.517	0.309	0.307	0.095	0.952
			$\hat{\beta}_C$	0.481	0.130	0.128	0.017	0.929	−0.475	0.217	0.221	0.048	0.958
			$\hat{\beta}_{opt}$	0.456	0.116	0.118	0.015	0.928	−0.464	0.202	0.204	0.042	0.943
150	250	15%	$\hat{\beta}_I$	0.478	0.192	0.188	0.037	0.944	−0.453	0.330	0.328	0.111	0.939
			$\hat{\beta}_P$	0.508	0.144	0.145	0.021	0.963	−0.512	0.245	0.246	0.060	0.946
			$\hat{\beta}_C$	0.500	0.119	0.119	0.014	0.949	−0.495	0.200	0.203	0.040	0.950
			$\hat{\beta}_{opt}$	0.493	0.116	0.115	0.014	0.943	−0.491	0.194	0.196	0.038	0.949
		30%	$\hat{\beta}_I$	0.422	0.187	0.182	0.041	0.918	−0.410	0.323	0.318	0.112	0.933

				β_1					B_2				
n_1	n_2	cr		**Est**	**SD**	**SE**	**MSE**	**CP**	**Est**	**SD**	**SE**	**MSE**	**CP**
			$\hat{\beta}_P$	0.504	0.156	0.158	0.024	0.959	−0.512	0.272	0.270	0.074	0.952
			$\hat{\beta}_C$	0.483	0.125	0.127	0.016	0.955	−0.482	0.215	0.218	0.047	0.950
			$\hat{\beta}_{opt}$	0.465	0.118	0.119	0.015	0.937	−0.470	0.203	0.205	0.042	0.959

TABLE 3

Distribution of risk factors by cohort.

Variable	**Incident only** **(n_1 = 153)**	**Prevalent only** **(n_2 = 77)**	**Combined cohorts** **(n = 230)**
APOE4			
Presence	39 (25%)	29 (38%)	68 (30%)
Absence	114 (75%)	48 (62%)	162 (70%)
EDCAT			
College and higher	134 (87%)	49 (64%)	183 (80%)
Others	19 (13%)	28 (36%)	47 (20%)

TABLE 4

Regression analysis under the proportional mean residual life model using data from incident cohort only, prevalent cohort only, and combined cohorts. Estimated parameter (Est) and standard error (SE).

	Incident only		**Prevalent only**		**Combined cohorts**	
	Est	**SE**	**Est**	**SE**	**Est**	**SE**
APOE4						
(presence=1, absence=0)	0.279	0.154	−0.278	0.245	0.148	0.131
EDCAT						
(college and higher=1, others=0)	−0.231	0.197	−0.470	0.253	−0.287	0.155