Estimating treatment effects in observational studies with both prevalent and incident cohorts

Jing Ning[1,*], **Chuan Hong**[2], **Liang Li**[1], **Xuelin Huang**[1], and **Yu Shen**[1]

[1]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston TX, 77030, USA

[2]Department of Biostatistics, Harvard School of Public Health, Boston MA, 02115, USA

Abstract

Registry databases are increasingly being used for comparative effectiveness research in cancer. Such databases reflect the real-world patient population and physician practice, and thus are natural sources for comparing multiple treatment scenarios and the associated long-term clinical outcomes. Registry databases usually include both incident and prevalent cohorts, which provide valuable complementary information for patients with more recent diagnoses in the incident cohort as well as patients with long-term follow-up data in the prevalent cohort. However, utilizing such data to derive valid inference poses two major challenges: the data from a prevalent cohort are not random samples of the target population, and there may be substantial differences in the baseline characteristics of patients between treatment arms, which influences the decisions about treatment selection in both cohorts. In this article, we extend propensity score methodology to observational studies that involve both prevalent and incident cohorts, and assess the effectiveness of radiation therapy in SEER-Medicare patients diagnosed with stage IV breast cancer. Specifically, we utilize the incident cohort to estimate the propensity for receiving radiation therapy, and then combine data from both the incident and prevalent cohorts to estimate the effect of radiation therapy by adjusting for the propensity scores in the model. We evaluate the proposed method with simulations. We demonstrate that the proposed propensity score method simultaneously removes sampling bias and selection bias under several assumptions.

Key words and phrases

Incident cohort; Prevalent cohort; Propensity score; Sampling bias; Selection bias

1. Introduction

The focus of comparative effectiveness research when using observational data is to assess the use of an intervention on a more general patient population outside of controlled clinical trials. Observational data cohorts such as national cancer registries and the Surveillance, Epidemiology and End Results (SEER)-Medicare databases are natural choices for conducting comparative effectiveness research. Due to advances in both breast cancer

*Author to whom correspondence may be addressed. jning@mdanderson.org.

screening programs and adjuvant systemic therapy, fewer breast cancer cases (5.6%, www.seer.cancer.gov) are metastatic (stage IV) at diagnosis than ever before, especially for women aged 65 years and older. There are only a few standards of care in the management of stage IV breast cancer (Cardoso et al., 2012) and the role of radiation therapy (RT) remains controversial for women aged 65 years and older who had stage IV breast cancer. On the other hand, a randomized clinical trial is unlikely to assess the role of RT for this cohort of patients.

Using the recent SEER-Medicare breast cancer database of 2007-2010, it is of great interest to evaluate the effectiveness of receiving RT versus not receiving RT in elderly patients with stage IV breast cancer based on the incident cohort. This cohort consists of patients who were diagnosed with breast cancer from 2007 on, and who died or were right-censored at last follow-up. An incident cohort study is not typically an efficient design for evaluating the treatment effect on the failure event of interest due to relatively short follow-up times and, in this example, the moderate size of the population of patients with stage IV breast cancer. Therefore, from the SEER-Medicare claims data, we also retrieved information on patients who were diagnosed with stage IV breast cancer prior to 2007 and who were alive in 2007. This represents a prevalent cohort consisting of patients who had already experienced the initiating event but had not experienced the failure event at the time of ascertainment. Figure 1 illustrates the difference in sampling scheme between a prevalent and an incident cohort study. By combining the data from the two cohorts with stage IV breast cancer, we may have a sufficiently large number of patients for a better assessment of the effect of RT.

Although we can take advantage of data from both prevalent and incident cohorts for more efficient statistical inference, we face two challenges. The first challenge is that the patients in the prevalent cohort often have longer survival times and the distribution of their covariates is not representative of the target population (i.e., incident cohort), as patients with longer survival times are preferentially sampled (Vardi, 1982; Wang et al., 1993; Gail and Benichou, 2000; Rothman et al., 2008; Zelen, 2006). The conventional estimator of the treatment effect from an outcome regression model using data from a prevalent cohort can be biased due to sampling bias related to the outcomes.

Moreover, the treatment of interest, such as RT, is not randomly assigned, and baseline patient characteristics are confounded with the treatment decision. Thus, the baseline characteristics are often imbalanced between patients who receive RT and those who do not, in either the incident or prevalent cohort. The propensity score method, introduced by Rubin (1973) and Rosenbaum and Rubin (1985), addressed this concern by reducing multiple baseline covariates to a single statistic that summarizes the collective information. Using this statistic, the imbalance in the baseline covariates between two treatment groups can be adjusted to reduce bias (D'Agostino, 1998). However, the existing methods are applicable to the incident cohort only. Prevalent cohorts suffer additional selection bias related to the length of the outcomes, since the patients in the cohort have all survived until their enrollment in the study, and those who did not survive until that time were automatically excluded.

Much progress has been made in the area of regression analysis to associate failure time and covariates given data from a prevalent cohort after adjusting for the sampling bias (Wang et al., 1993; Shen et al., 2009; Qin and Shen, 2010; Ning et al., 2011). However, little has been done regarding propensity score analysis using both prevalent and incident cohorts. Cheng and Wang (2012) and Chan (2013) developed statistical methods for propensity score analysis using prevalent cohort data only, in which some restrictive model assumptions had to be imposed.

Our strategy is to use the incident cohort to estimate the propensity for receiving RT and then use both cohorts to estimate the effectiveness of receiving RT versus not receiving RT in elderly patients with stage IV breast cancer, by using an outcome regression model (e.g., Cox model) involving the propensity scores. The sampling bias in the prevalent cohort has been corrected in the estimation of the treatment effect.

The remainder of this article is organized as follows. In Section 2, we present the proposed estimating equations and inference procedure to simultaneously estimate the regression parameters in models for the propensity score and survival outcome. In Section 3, we develop procedures to check whether the propensity scores estimated from the incident cohort can be applied to the prevalent cohort to balance the covariate distributions between the treatment arms. In Section 4, we report the simulation studies to assess the finite sample performance of the proposed method. We apply our method to the SEER-Medicare data in Section 5, and provide concluding remarks in Section 6. We provide details for the proofs of the asymptotic properties in the Appendix.

2. Model and Estimation

2.1. Notations and Model

We consider data from the incident cohort to represent the target population. Let T, Z and $\mathbf{X}$ respectively represent the failure time, treatment indicator (RT versus no RT), and a $q \times 1$ vector of covariates in the incident cohort. Define the censoring time and censoring indicator to be C and $\delta = I(T < C)$, respectively. Then the observed data from the incident cohort are $\{Y_i, Z_i, \mathbf{X}_i, \delta_i; i = 1, \cdots, m\}$, where Y_i is $\min(T_i, C_i)$, m is the sample size of the incident cohort, and $I(.)$ is the indicator function.

In the prevalent cohort, patients with the diagnosis of stage IV breast cancer prior to 2007, the time of enrollment to the study cohort, who were still at risk for the failure event (e.g., breast cancer-specific death) in 2007 were included (e.g., subjects 5 and 6 in Figure 1). The observed truncation time W is the duration from the initial diagnosis of the disease to 2007. Thus, the prevalent cohort excludes patients with $T < W$ (e.g., subject 4 in Figure 1).

Data from the prevalent cohort represent a valuable augmentation to the incident cohort; the observed survival times from the prevalent cohort tend to be longer than those in the incident cohort. Moreover, the distributions of the baseline variables can differ between the incident and prevalent cohorts. We need a slightly different notation to define the observed data in the prevalent cohort. Let $\tilde{T}$, $\tilde{W}$, $\tilde{Z}$ and $\tilde{\mathbf{X}}$ respectively be the failure time, truncation time, treatment indicator, and a vector of the covariates. Denote the residual survival time by $\tilde{V}$, so

that $\tilde{T} = \tilde{W} + \tilde{V}$. The observed data are recorded as $\{ \tilde{Y}_i, \tilde{W}_i, \tilde{Z}_i, \tilde{\mathbf{X}}_i, \tilde{\delta}_i; i = 1, \cdots, n \}$, where $\tilde{Y}_i = \min\{ \tilde{W}_i + \tilde{V}_i, \tilde{C}_i \}$, $\tilde{\delta}_i = I(\tilde{W}_i + \tilde{V}_i < \tilde{C}_i)$ and n is the sample size of the prevalent cohort. The propensity score method is used to mitigate the treatment selection bias arising from non-random treatment assignment in observational studies (Rosenbaum and Rubin, 1983; Kang and Schafer, 2007; Austin, 2008). We assume that the propensity score, i.e., the probability of receiving the treatment, is associated with the baseline patient characteristics via a logistic regression model,

$$e(\mathbf{X};\boldsymbol{\gamma}) = P(Z=1|\mathbf{X};\boldsymbol{\gamma}) = \frac{\exp\{\boldsymbol{\gamma}^{\top}(1,\mathbf{X})\}}{1+\exp\{\boldsymbol{\gamma}^{\top}(1,\mathbf{X})\}}, \quad (1)$$

where $\boldsymbol{\gamma}$ is a vector of regression coefficients, and the interaction terms of $\mathbf{X}$ may be included to increase the flexibility of the propensity score model. As shown by D'Agostino (1998), Rosenbaum and Rubin (1985) and Rubin (1979), using the propensity score as a covariate in an outcome regression analysis can effectively reduce the bias in the estimation of the treatment effect in observational studies. To determine the association between the survival outcome T and RT treatment, the Cox proportional hazards model is used to adjust for the effects of baseline covariates by including the propensity score as a covariate:

$$\lambda(t|Z,\mathbf{X}) = \lambda_0(t)\exp[\alpha Z + g\{e(\mathbf{X},\boldsymbol{\gamma});\boldsymbol{\beta}\}], \quad (2)$$

where $\lambda_0(.)$ is an unspecified baseline function, α is the log-hazard ratio of the RT treatment after the propensity score adjustment, $\boldsymbol{\beta}$ is a vector of parameters in $g(.)$ and $g(.)$ is a flexible and differentiable function.

One major advantage of using the propensity score as a covariate in the outcome regression is to avoid over-parameterizing or mis-specifying the model with a large number of covariates. By collapsing all covariates into a single propensity score and incorporating the plausible nonlinear and interaction terms of the covariates into the propensity score, one can greatly reduce the possibility of outcome model misidentification (D'Agostino, 1998). Regression adjustment with the propensity score was used by Cheng and Wang (2012) to analyze prevalent survival data. Although the propensity score analysis introduces an additional assumption of a propensity score model, that assumption can be reinforced by assessing the balance in the covariate distributions between the two groups. The propensity score model can be reformulated until adequate balance is achieved. From this perspective, the propensity score analysis involves weaker assumptions than direct regression on the outcomes using a large number of covariates, and the estimation of α is more robust than an estimation that is conditional on all the covariates.

2.2. Estimating Equation Methods

Without loss of generality, we use a simple linear form of $g(.)$ for illustration, $g\{e(\mathbf{X}; \boldsymbol{\gamma}); \boldsymbol{\beta}\}$ $= \beta e(\mathbf{X}; \boldsymbol{\gamma})$. Using the data from the incident cohort, we estimate the propensity score by solving the score equations under the logistic regression model,

$$U_1(\boldsymbol{\gamma}) = \sum_{i=1}^{m} (1, \mathbf{X}_i)^\top \left[Z_i - \frac{\exp\{\boldsymbol{\gamma}^\top (1, \mathbf{X}_i)\}}{1 + \exp\{\boldsymbol{\gamma}^\top (1, \mathbf{X}_i)\}} \right] = 0 \tag{3}$$

The score equation under the Cox model using the incident cohort is

$$\sum_{i=1}^{m} \delta_i \left[\{Z_i, e(\mathbf{X}_i; \boldsymbol{\gamma})\}^\top - \frac{\sum_{j \in R_1(Y_i)} \{Z_j, e(\mathbf{X}_j; \boldsymbol{\gamma})\}^\top \exp\{\alpha Z_j + \beta e(\mathbf{X}_j; \boldsymbol{\gamma})\}}{\sum_{j \in R_1(Y_i)} \exp\{\alpha Z_j + \beta e(\mathbf{X}_j; \boldsymbol{\gamma})\}} \right] = 0, \tag{4}$$

where $R_1(t) = \{j : Y_j \geq t\}$ is the at-risk function for the incident cohort. We estimate $\boldsymbol{\gamma}$, α and β by solving (3) and (4) when using the incident cohort only.

Next, we describe how to combine the data from the prevalent cohort to improve statistical efficiency. In our motivating example, the incident cohort and prevalent cohort are derived from the same target population but with different sampling schema. Thereby, the propensity scores estimated from the incident cohort can be applied to the prevalent cohort to balance the baseline variables. The Cox model structure assumed for the incident cohort can be different from the one for the observed data from the prevalent cohort. Therefore, we need to adjust for biased sampling due to left-truncation when modeling the treatment effect on the survival outcome by using a modified at-risk function (Lai and Ying, 1991; Wang et al., 1993). We extend score equation (4) to incorporate the information from the prevalent cohort:

$$\begin{aligned}
U_2(&\boldsymbol{\gamma}, \alpha, \beta) \\
&= \sum_{i=1}^{m} \delta_i \left[\{Z_i, e(\mathbf{X}_i; \boldsymbol{\gamma})\}^\top \right. \\
&\left. - \frac{\sum_{j \in R_1(Y_i)} \{Z_j, e(\mathbf{X}_j; \boldsymbol{\gamma})\}^\top \exp\{\alpha Z_j + \beta e(\mathbf{X}_j; \boldsymbol{\gamma})\}}{\sum_{j \in R_1(Y_i)} \exp\{\alpha Z_j + \beta e(\mathbf{X}_j; \boldsymbol{\gamma})\}} + \sum_{i=1}^{n} \tilde{\delta}_i \left[\{\tilde{Z}_i, e(\tilde{\mathbf{X}}_i; \boldsymbol{\gamma})\}^\top \right. \right. \\
&\left. \left. - \frac{\sum_{j \in R_2(\tilde{Y}_i)} \{\tilde{Z}_j, e(\tilde{\mathbf{X}}_j; \boldsymbol{\gamma})\}^\top \exp\{\alpha \tilde{Z}_j + \beta e(\tilde{\mathbf{X}}_j; \boldsymbol{\gamma})\}}{\sum_{j \in R_2(\tilde{Y}_i)} \exp\{\alpha \tilde{Z}_j + \beta e(\tilde{\mathbf{X}}_j; \boldsymbol{\gamma})\}} \right] = 0,
\end{aligned} \tag{5}$$

where $R_2(t) = \{j : \tilde{W}_j \leq t \leq \tilde{Y}_j\}$ is the modified at-risk function for the left-truncated and right-censored data for the prevalent cohort. Then, estimating equations $U_1(\boldsymbol{\gamma})$ and $U_2(\boldsymbol{\gamma}, \alpha, \beta)$ are simultaneously solved to estimate the parameters of interest, $\boldsymbol{\gamma}$, β and α. The first

component of $U_2(.)$ uses data from the incident cohort, which often has shorter follow-up outcomes than the prevalence cohort but is not subject to biased sampling; whereas the second component of $U_2(.)$ uses data from the prevalent cohort, which often has longer follow-up outcomes but requires an adjustment of biased sampling.

Define $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\gamma}}^\top, \hat{a}, \hat{\beta})^\top$ as the solution of $\boldsymbol{U}(\boldsymbol{\theta}) = (U_1(\boldsymbol{\gamma})^\top, U_2(\boldsymbol{\theta})^\top)^\top = 0$ and $\boldsymbol{\theta}_0$ as the true value of $\boldsymbol{\theta}$. We show in the Appendix that $\boldsymbol{U}(\boldsymbol{\theta}) = 0$ asymptotically has a unique solution $\hat{\boldsymbol{\theta}}$, and then derive the consistency of $\hat{\boldsymbol{\theta}}$. Moreover, based on Taylor expansions of $\boldsymbol{U}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$, the central limit theorem, and the martingale central limit theorem, we can prove that $\hat{\boldsymbol{\theta}}$ is asymptotically normally distributed under certain regularity conditions. We summarize these results in the following theorem, whose proof is presented in the Appendix.

Theorem 1: Under the logistic regression model for the propensity score and the Cox model for the survival time and regularity conditions [C.1-C.6] listed in the Appendix, the estimator $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}_0$ in probability. Moreover, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution to a normal distribution with mean 0 and covariance matrix defined in the Appendix.

2.2.1. Special Case: Length-biased Data—If the onset of the disease follows a Poisson process, the observed survival data from the prevalent cohort are considered to be length-biased data (Wang et al., 1993). The aforementioned set of estimating equations can be modified to increase statistical efficiency by incorporating the distributional information of the truncation time. Using the weight in a manner similar to the approach of Shen et al. (2009) and Qin and Shen (2010), we restructure unbiased estimating equations for estimating $\boldsymbol{\gamma}$ by using the baseline covariate data from the prevalent cohort: $U_{LB1}(\boldsymbol{\gamma}) = 0$, where

$$U_{LB1}(\boldsymbol{\gamma}) = \sum_{i=1}^{m}(1, \mathbf{X}_i)^\top \left\{ Z_i - \frac{\exp\{\boldsymbol{\gamma}^\top(1, \mathbf{X}_i)\}}{1+\exp\{\boldsymbol{\gamma}^\top(1, \mathbf{X}_i)\}} \right\} + \sum_{j=1}^{n} \frac{\tilde{\delta}_j(1, \tilde{\mathbf{X}}_j)^\top}{w(\tilde{Y}_j)} \left\{ \tilde{Z}_i - \frac{\exp\{\boldsymbol{\gamma}^\top(1, \tilde{\mathbf{X}}_i)\}}{1+\exp\{\boldsymbol{\gamma}^\top(1, \tilde{\mathbf{X}}_i)\}} \right\},$$

(6)

$w(t) = \int_0^t S_c(u)\mathrm{d}u$, and $S_c(.)$ is the survival distribution for the residual censoring time $\tilde{C} - \tilde{W}$ which can be consistently estimated by the Kaplan-Meier estimator. The second component of (6) is similar to the first one, but is inversely adjusted by the weight $w(\tilde{Y}_j)$ for data from the prevalent cohort (Ertefaie, 2014).

Similarly, the second set of estimating equations under the Cox model can be reconstructed to efficiently utilize the length-biased data: $U_{LB2}(\boldsymbol{\gamma}, a, \beta) = 0$, where

$$U_{LB2}(\boldsymbol{\gamma}, \alpha, \beta)$$

$$= \sum_{i=1}^{m} \delta_i \left[\{Z_i, e(\mathbf{X}_i;\boldsymbol{\gamma})\}^\top \right.$$

$$- \frac{\sum_{j \in R_1(Y_i)} \{Z_j, e(\mathbf{X}_j;\boldsymbol{\gamma})\}^\top \exp\{\alpha Z_j + \beta e(\mathbf{X}_j;\boldsymbol{\gamma})\}}{\sum_{j \in R_1(Y_i)} \exp\{\alpha Z_j + \beta e(\mathbf{X}_j;\boldsymbol{\gamma})\}} + \sum_{i=1}^{n} \tilde{\delta}_i [\{\tilde{Z}_i, e(\tilde{\mathbf{X}}_i, \boldsymbol{\gamma})\}^\top$$

$$\left. - \frac{\sum_{j \in R_1(\tilde{Y}_i)} \tilde{\delta}_j \{w(\tilde{Y}_j)\}^{-1} \{\tilde{Z}_j, e(\tilde{\mathbf{X}}_j;\boldsymbol{\gamma})\}^\top \exp\{\alpha \tilde{Z}_j + \beta e(\tilde{\mathbf{X}}_j;\boldsymbol{\gamma})\}}{\sum_{j \in R_1(\tilde{Y}_i)} \tilde{\delta}_j \{w(\tilde{Y}_j)\}^{-1} \exp\{\alpha \tilde{Z}_j + \beta e(\tilde{\mathbf{X}}_j;\boldsymbol{\gamma})\}} \right].$$

(7)

In the second component of (7), the subjects in the risk sets are weighted to adjust for the length-biased sampling constraint. By using derivations similar to those of Qin and Shen (2010), (7) has mean zero and can be combined with (6) for estimating θ. The consistency and weak convergence of the resultant estimators can be established using techniques similar to those for Theorem 1.

3. Testing the Balancing Property

Unlike the motivating example, the incident cohort and prevalent cohort may not be generated from the same target population in some applications. In this case, we need to check the assumptions for the proposed method to determine whether the propensity scores estimated from the incident cohort can balance the covariate distributions of the two treatment arms in the prevalent cohort. This test is conducted under the following modeling assumption:

$$E(\tilde{\mathbf{X}}_i) - \boldsymbol{\xi}\mu\{e(\tilde{X}_i;\hat{\boldsymbol{\gamma}})\} + \boldsymbol{\zeta}\hat{Z}_i,$$

where $\mu(.)$ is a pre-specified flexible function (e.g., fractional polynomial function or spline function), and $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ are vectors with the same dimension of $\tilde{\mathbf{X}}_i$.

Under this model specification, testing the null hypothesis of the covariate being independent of the treatment assignment is equivalent to testing for $H_0 : \boldsymbol{\zeta} = 0$ vs $H_1 : \boldsymbol{\zeta} \neq 0$. This test can also be viewed as a test for the specification of the propensity score model (1) (Imai et al., 2008; Li and Greene, 2013). Rejection of the null hypothesis indicates that the propensity score model derived from the incident cohort is not applicable to the prevalent cohort.

Under the assumption that the initial event has stationary incidence, we construct an unbiased estimating equation for estimating the unknown parameters and making inference on $\boldsymbol{\zeta}$:

$$\sum_{i=1}^{n} \frac{\tilde{\delta}_i}{w(\tilde{Y}_i)} \left\{ \tilde{\mathbf{X}}_i - \boldsymbol{\xi}\mu\{e(\tilde{\mathbf{X}}_i;\hat{\boldsymbol{\gamma}})\} - \boldsymbol{\zeta}\tilde{Z}_i \right\} = 0 \tag{8}$$

The unknown weight function $w(\tilde{Y}_i)$ involved in (8) can be readily estimated by the Kaplan-Meier estimator. In the Appendix, we have derived the asymptotic normality of estimator $\hat{\boldsymbol{\zeta}}$, the solution of (8). We propose a Wald-type test statistic to test for H_0 by $T_{test} = \hat{\boldsymbol{\zeta}}^\top \hat{V}^{-1} \hat{\boldsymbol{\zeta}}$, where $\hat{V}$ is the variance-covariance matrix of $\hat{\boldsymbol{\zeta}}$. The following theorem summarizes the limiting distribution of the test statistics. A sketch of the proof is provided in the Appendix.

Theorem 2: Under regularity conditions [C.1-C.7] listed in the Appendix and the null hypothesis, T_{test} converges weakly to a chi-squared distribution with a degree of freedom of q.

The weight function in (8) can be generalized to accommodate general left-truncated data for which the stationarity assumption is not satisfied. If the truncation time follows a density function $h_{\boldsymbol{\theta}_A}$ with a finite dimensional vector of parameters $\boldsymbol{\theta}_A$, the generalized weight function is $w(\tilde{Y}_\top) = \int_0^{\tilde{Y}_i} h_{\boldsymbol{\theta}_A}(a) S_c(Y_i - a)\mathrm{d}a$. Similarly, the weight can be estimated consistently by plugging in consistent estimators for the censoring survival function and $\boldsymbol{\theta}_A$.

4. Simulation Study

We conducted simulation studies to evaluate the finite sample performance of the proposed methods. We simulated two cohorts from the same target population: $m = 200$, 400 or 600 subjects in the incident cohort and $n = 200$ patients in the prevalent cohort. For both cohorts, we generated the treatment assignment Z_i via the propensity score function

$$e(\mathbf{X}_i, \boldsymbol{\gamma}) = \exp(\boldsymbol{\gamma}^\top (1, \mathbf{X}_i))/\{1 + \exp(\boldsymbol{\gamma}^\top (1, \mathbf{X}_i))\},$$

where the baseline covariates $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3})$ followed the standard normal distribution and $\boldsymbol{\gamma} = (0.1, 0.4, 0.4, 0.4)$. Given $(Z_i, \mathbf{X_i})$, the failure time T_i was generated from an exponential distribution with a rate of $\exp\{0.5Z_i + e(\mathbf{X}_i; \boldsymbol{\gamma})\}$. The censoring times were independently generated from the uniform distribution $[0, \tau_c]$, where $\tau_c = 0.52$ in Scenario I and $\tau_c = 1.3$ in Scenario II. The sampling constraint in the prevalent cohort was introduced by a uniform truncation time; only patients satisfying $W_i < T_i$ were included in the prevalent cohort. We considered different combinations of sample sizes and censoring rates to cover a wide range of scenarios. We used estimating equations $U_1(\boldsymbol{\gamma})$ and $U_2(\boldsymbol{\gamma}, a, \beta)$ to fit the propensity score and survival outcome, where $g\{e(\mathbf{X};\boldsymbol{\gamma}); \boldsymbol{\beta}\} = \beta e(\mathbf{X}; \boldsymbol{\gamma})$.

Table 1 summarizes the empirical biases, empirical standard errors, average asymptotic standard errors and coverage probabilities of the 95% confidence intervals based on 1, 000 simulations. As shown, the parameters in both the propensity score model and the Cox model were estimated reliably by the proposed method. The empirical biases were close to zero, and the model-based standard errors were close to their empirical values. Furthermore,

the coverage probabilities of the 95% confidence intervals were close to the nominal level. As expected, the standard errors decreased with increasing sample sizes, and the standard errors of parameter estimates in the Cox model increased with increasing rates of censoring. For comparison (but not reported here), we also calculated the standard errors of $\hat{a}$ and $\hat{\beta}$ by ignoring the additional variation due to the estimated propensity score. We found that the standard errors were underestimated and the associated coverage probabilities were much smaller than the nominal value. For example, in the scenario with a sample size of 200 from each cohort, the estimated standard error of $\hat{\beta}$ was 0.409, which is substantially lower than its empirical standard deviation of 0.576 in Table 1. We compared the small sample performance of the estimators from the estimation equations $U_1(\boldsymbol{\gamma})$ and $U_2(\boldsymbol{\gamma}, a, \beta)$, and the estimation equations $U_{LB1}(\boldsymbol{\gamma})$ and $U_{LB2}(\boldsymbol{\gamma}, a, \beta)$. As expected, by incorporating the distribution information of the truncation time, the estimators obtained from estimation equations $U_{LB1}(\boldsymbol{\gamma})$ and $U_{LB2}(\boldsymbol{\gamma}, a, \beta)$ are more efficient than those from estimation equations $U_1(\boldsymbol{\gamma})$ and $U_2(\boldsymbol{\gamma}, a, \beta)$, which are for general left-truncated data. More details are provided in the Supplementary Materials.

We conducted some sensitivity studies to evaluate robustness of the proposed method with violations of model assumptions. The simulation results and findings are summarized in the Supplementary Materials.

Table 2 lists the simulation results of two naive methods: one uses the incident cohort data only, the other uses the prevalent cohort data only without adjusting for the sampling bias. As expected, the first naive method was valid, with small empirical biases and coverage probabilities close to the nominal value, but was much less efficient than the proposed method, which combined data from the two cohorts. In contrast, the second naive method produced biased estimates and misleading inference by ignoring that neither the baseline covariate distribution nor the survival outcomes were representative of the target population.

The second set of simulations evaluates the finite sample performance (size and power) of the proposed test for determining the covariate balance in the prevalent cohort under various scenarios. We used 5000 replications of the tests to calculate the size and power, and let $\mu\{e(\tilde{\mathbf{X}}_i; \hat{\boldsymbol{\gamma}})\} = e(\tilde{\mathbf{X}}_i; \hat{\boldsymbol{\gamma}})$. We generated data for the incident cohort from the same model described in the first set of simulations. Under the null hypothesis, the propensity score of the prevalent cohort is the same as that of the incident cohort, $e(\mathbf{x}_i) = P(Z_i|\mathbf{X}_i) = \text{expit}(0.1 + 0.4X_{i1} + 0.4X_{i2} + 0.4X_{i3})$, where $\text{expit}(.) = \exp(.)/\{1 + \exp(.)\}$. Under the alternative (power I), we set the propensity score of the prevalent cohort as $e(\mathbf{x}_i) = P(Z_i|\mathbf{X}_i) = \text{expit}(0.1 + 0.4X_{i1}^2 + 0.4X_{i2}^2 + 0.4X_{i3}^2)$, which has a different form than the propensity score function of the incident cohort. For another alternative (power II), we chose the propensity score of the prevalent cohort as $e(\mathbf{x}_i) = P(Z_i|\mathbf{X}_i) = \text{expit}(0.1 + 0.8X_{i1} + 1.2X_{i2} + 1.6X_{i3})$, which has the same function form but different coefficients from that in the incident cohort.

Table 3 lists the rates of rejecting the null hypothesis at a significance level of 0.05. The type I error rates were well maintained in the null scenario, especially for larger sample sizes. The power of the test increased with increasing sample sizes and did not change with the censoring rate. However, the proposed test performed differently for the two alternative

settings. The test was very powerful for detecting different function forms between the two propensity scores (power scenario I); but, it had only mild power to detect the difference when the propensity scores shared the same model form but had different regression coefficients (power scenario II). Because the proposed test has relatively low power under power scenario II, we applied the proposed estimation method to the data generated from power scenario II when the difference in propensity score models was ignored. The resulting treatment estimators were reasonably robust to such model misspecification, and did not have substantial biases; for example, the bias was 0.05 (true value=0.5) and the associated coverage probability was 0.933 when $m = 400$ and $n = 200$.

5. Data Application

We used data from the SEER-Medicare database, which links the National Cancer Institutes SEER registry with Medicare claims and enrollment files. The SEER-Medicare database connects 94% of patients aged 65 years or older. Using this population-based database, we identified a prevalent cohort and an incident cohort to evaluate the effect of RT in patients of age 66 who were diagnosed with stage IV breast cancer as the first primary cancer. The incident cohort consisted of 1867 patients who were diagnosed with breast cancer after 2007, and the prevalent cohort included 1106 patients who were diagnosed with breast cancer prior to 2007 and were alive in 2007. Patients in both cohorts were followed until death or the last observation time as of December 31, 2010.

The clinical question of interest was whether RT had any benefit on the overall survival time measured from the date of breast cancer diagnosis. We extracted each patient's baseline covariates, including comorbidity, race, region of residence, age at diagnosis, estrogen-receptor status, progesterone-receptor status, and tumour size. We used Medicare claims to identify RT received within 12 months of breast cancer diagnosis. Table 4 summarizes the patient characteristics by cohort. Using the Kaplan-Meier method, the estimated median survival times for the prevalent cohort were much longer than those for the incident cohort, which suggested that the observed outcome data from the prevalent cohort were sampled with bias. We used the t-test or chi-square test to detect the difference in the covariate distributions between the two cohorts. Except for race, all covariates had different distributions between the two cohorts, which implied that the observed covariates in the prevalent cohort likely did not represent the target population.

For patients older than 65 years who had stage IV breast cancer, those who received RT might differ from those who did not receive RT. For example, it is likely that patients who received RT were healthier than patients who did not, regardless of whether there was any clinical benefit of RT in that age group. Hence, we included the propensity score in the Cox model to balance the covariate distributions between the RT and no-RT patient groups. We first used the baseline covariates from the 1867 patients in the incident cohort to estimate the propensity scores of a patient receiving RT, and then used the combined cohort to fit the Cox model while controlling for and not controlling for the propensity score. As seen in Table 5, a much larger beneficial effect of RT (hazard ratio=0.66) was found in the Cox model when we ignored the covariate imbalance between the RT and no-RT groups, which was likely to

be overestimated. When controlling for the covariate imbalance by the propensity score, the hazard ratio of receiving RT was 0.76, with a standard error of 0.05.

Our next step was to combine the prevalent and incident cohorts to gain statistical efficiency. Using the test in Section 3, we checked whether the estimated propensity scores from the incident cohort can be applied to the prevalent cohort to balance the covariate distribution between the RT and no-RT groups. The p-value was 0.99, which suggested that the RT and no-RT groups within the prevalent cohort had comparable baseline covariate distributions after applying the propensity score from the incident cohort. We then included the prevalent data to fit the Cox models after adjusting for baseline covariates and sampling bias in the prevalent cohort. The results indicated that the use of RT had a significant benefit for the overall survival time (hazard ratio=0.75; p-value< 0.001), and the combined data yielded a similar RT effect, but with a smaller standard error (0.03 vs 0.05) compared to that from the incident cohort only.

6. Discussion

Propensity score analysis is a valuable and convenient approach to the analysis of observational data. If data come from a prevalent cohort study, the propensity score analysis encounters significant difficulty because of sampling bias. However, prevalent cohorts contain valuable information on clinical outcomes, usually with longer periods of follow-up compared with incident cohorts. In this article, we propose a propensity score method based on the joint analysis of the incident and prevalent cohorts, which allows for adjustment for both the covariate imbalance between treatment arms and the sampling bias in the prevalent cohort. We assume both cohorts come from the same target population so that the joint analysis of both cohorts would provide more efficient results than the analysis of the incident cohort alone.

We develop a test procedure to check whether the propensity score model estimated with the incident cohort can balance the covariates between treatment groups in the prevalent cohort. Once the balance property of the prevalent cohort is confirmed by the test, we include the estimated propensity score in the Cox model and propose estimating methods, in which the modified at-risk function or the inverse weighting method is adopted to handle the biased sampling issue in the prevalent cohort. Given additional distribution assumptions on the truncation time, the data from the prevalent cohort can be utilized to estimate the propensity score and improve statistical efficiency.

It has been argued that a specification test can be used to check the adequacy of the propensity score model (Li and Greene, 2013; Imai et al., 2008; Hansen and Bowers, 2008). In this article, we present an implementation of this idea in the context of the joint analysis of prevalent and incident cohorts. For the proposed test, we restrict our attention to the setting where the censoring time and truncation time are independent of the covariates. To accommodate the possibility of covariate-dependent censoring and covariate-dependent truncation, we can adopt regression analysis or a local Kaplan-Meier estimator (González-Manteiga and Cadarso-Suarez, 1994) to estimate the covariate-specific distributions, denoted as $h_{\boldsymbol{\theta}_A}(a|z, \mathbf{x})$ and $S_c(t|z, \mathbf{x})$. The test procedure can be generalized to incorporate covariate-

dependent censoring and covariate-dependent truncation, by replacing the weighting

function in (8) by $w(\tilde{Y}_i|Z_i,\mathbf{x}_i)=\int_0^{\tilde{Y}_i} h_{\boldsymbol{\theta}_A}(a|z_i,\mathbf{x}_i)S_c(Y_i-a|z_i,\mathbf{x}_i)\mathrm{d}a$. After the replacement we plug the estimated covariate-specific distributions in the implementation of the test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Appendix

A.1. Regularity conditions

> C.1) Given the observable covariate X, potential outcomes are independent of the treatment assignment.
>
> C.2) The propensity score $e(x) \in (0, 1)$, such that subjects with the same covariate values have a positive probability of being assigned to either of two treatment arms.
>
> C.3) The parameter space of $\boldsymbol{\theta}$ is a compact subset of $\mathbb{R}^{q+3}$, and the true parameter value $\boldsymbol{\theta}_0$ is in the interior of the parameter space.
>
> C.4) $\mathbf{X}$ is a $q \times 1$ vector of bounded covariates, not contained in a $(q-1)$-dimensional hyperplane.
>
> C.5) The censoring time C is not degenerate at 0 given any (z, x), and $P(T > C|Z, X) > 0$.
>
> C.6) The conditional survival function of $\tilde{C}$ satisfies $t_0 \quad \sup\{t: \tilde{S}_c(t|\tilde{Z}, \hat{X}) > 0\} \quad t_1$ uniformly in $(\tilde{Z}, \tilde{X})$ for some positive constants t_0 and t_1. Furthermore, $P(\tilde{V} > \tilde{C}|\tilde{Z}, \tilde{X}) > 0$.
>
> C.7) $\int_0^{\tau_0} \left[\{\int_t^{\tau_0} S_c(u)\mathrm{d}u\}^2 / \{S_c(t)^2 S_v(t)\} \right] dS_c(t) < \infty$, where $S_v(.)$ is the survival function for the residual failure time.

Condition C.1, referred to as ignorability assumption, implies that all variables that influence treatment assignment and potential outcomes are observed. Condition C.2 rules out the phenomenon of perfect predictability of the treatment received given X. Condition C.3 ensures the compactness of the parameter space, and Condition C.4 is for model identifiability in the regression analysis. Conditions C.5 and C.6 state assumptions for the censoring time of the incident cohort and the residual censoring time of the prevalent cohort such that there is a probability of observing uncensored observations in both cohorts. Condition C.7 is a required assumption for the weak convergence of the Kaplan-Meier estimator of the residue censoring.

A.2. Proof of Theorem 1

Note that the estimating equation $U_1(\boldsymbol{\gamma}) = 0$ is the score estimating equation under the logistic regression model. Under regularity conditions C.3 and C.4, the maximum likelihood estimator $\hat{\boldsymbol{\gamma}}$ is consistent.

Under the Cox model, the estimating equation $U_2(\theta) = 0$ is the sum of the partial score equation for the data from the incident cohort and the conditional score equation for the data from the prevalent cohort. Hence the solution to $U_2(\theta) = 0$, denoted as $\hat{\theta}$, is the maximum of the likelihood $L(\theta) = L_1(\theta) \times L_2(\theta)$, where $L_1(\theta)$ is the partial likelihood of the data from the incident cohort and $L_2(\theta)$ is the conditional likelihood for the data from the prevalent cohort. For simplicity of notation, let $\theta_0 = 0$. For any $\delta = 0$, since the intersection of the parameter space of θ and the closure of a δ-neighborhood of origin is closed, log $L(\theta)$ has a local maximum on this set. If we can show that the maximum lies in the parameter space of θ at a distance from the origin less than δ in probability, then the existence and consistency of $\hat{\theta}$ will follow immediately. This can be shown by proving that log $L(\theta) < $ log $L(0)$ with probability tending toward 1 for all θ in its parameter space that are at a distance δ from the origin. Such a fact can be established by using the Taylor expansion of log log $L(\theta)$ around 0. For simplicity, we assume θ is one-dimensional, then we have

$$\frac{1}{n}\{L(\delta) - L(0)\} = \frac{1}{n}\frac{\partial \log L(0)}{\partial \theta} + \frac{1}{2n}\frac{\partial^2 \log L(\delta^*)}{\partial \theta^2}\delta^2,$$

where δ^* is a point between the origin and δ. Note that when n and m go to infinity,

$$\frac{1}{n}\frac{\partial \log L(0)}{\partial \theta} \to^p 0 \text{ and } \frac{1}{2n}\frac{\partial^2 \log L(\delta^*)}{\partial \theta^2} \to^p a < 0.$$

Then we have $\{L(\delta) - L(0)\} < 0$ in probability. Consistency follows.

For the asymptotic normality, apply the Taylor expansion of $U(\theta)$ around θ_0:

$$\sqrt{N}(\hat{\theta} - \theta_0) = \left\{-\frac{1}{N}\frac{\partial U(\theta_0)}{\partial \theta}\right\}^{-1}\frac{1}{\sqrt{N}}U(\theta_0) + o(1),$$

where $N = n + m$. By the strong law of large numbers, $\dfrac{\partial U(\hat{\theta})}{\partial \theta}/N$ converges to

$$A = [\rho E\{\partial U_{i1}(\theta)/\partial \theta\}, \rho E\{\partial U_{i21}(\theta)/\partial \theta\} + (1 - \rho)E\{\partial U_{i22}(\theta)/\partial \theta\}] \quad (9)$$

almost surely, where $\rho = \lim_{n\to\infty, m\to\infty} m/(m + n)$,

$$U_{i1}(\boldsymbol{\theta})=(1,\mathbf{X}_i)^\top\left\{Z_i-\frac{\exp\{\boldsymbol{\gamma}^\top(1,\mathbf{X}_i)\}}{1+\exp\{\boldsymbol{\gamma}^\top(1,\mathbf{X}_i)\}}\right\}U_{i21}(\boldsymbol{\theta}($$

$$=\delta_i\left[\{Z_i,e(\mathbf{X}_i;\boldsymbol{\gamma})\}^\top\right.$$

$$-\frac{\sum_{j\in R_1(Y_i)}\{Z_j,e(\mathbf{X}_j;\boldsymbol{\gamma})\}^\top\exp\{\alpha Z_j+\beta e(\mathbf{X}_j;\boldsymbol{\gamma})\}}{\sum_{j\in R_1(Y_i)}\exp\{\alpha Z_j+\beta e(\mathbf{X}_j;\boldsymbol{\gamma})\}}U_{i22}(\boldsymbol{\theta}(=\tilde{\delta}_i\left[\{\tilde{Z}_i,e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma})\}^\top\right.$$

$$\left.-\frac{\sum_{j\in R_2(\tilde{Y}_i)}\{\tilde{Z}_j,e(\tilde{\mathbf{X}}_j;\boldsymbol{\gamma})\}^\top\exp\{\alpha\tilde{Z}_j+\beta e(\tilde{\mathbf{X}}_j;\boldsymbol{\gamma})\}}{\sum_{j\in R_2(\tilde{Y}_i)}\exp\{\alpha\tilde{Z}_j+\beta e(\tilde{\mathbf{X}}_j;\boldsymbol{\gamma})\}}\right].$$

The matrix $\boldsymbol{A}$ can be empirically estimated by definition. Applying the classical central limit theorem and martingale central limit theorem, as $n\to\infty$ and $m\to\infty$, we have

$$\boldsymbol{U}(\boldsymbol{\theta}_0)/\sqrt{N}\to^d N(\boldsymbol{0},\textstyle\sum),\quad(10)$$

where $\sum=\begin{pmatrix}\boldsymbol{\sigma}_{11}^2 & 0\\ 0 & \boldsymbol{\sigma}_{22}^2\end{pmatrix}$, $\boldsymbol{\sigma}_{11}^2=\rho E\{\boldsymbol{U}_{i1}(\boldsymbol{\theta})\}^{\otimes2}$, and

$\boldsymbol{\sigma}_{22}^2=\rho\mathrm{var}\{\boldsymbol{U}_{i21}(\boldsymbol{\theta})\}+(1-\rho)\mathrm{var}\{\boldsymbol{U}_{i22}(\boldsymbol{\theta})\}$. The covariance matrix that corresponds to estimating equation $\boldsymbol{U}_1(\boldsymbol{\theta}_0)$ can be consistently estimated by

$$\hat{\sigma}_{11}^2=\rho\frac{1}{m}\sum_{i=1}^m\left\{\boldsymbol{U}_{i1}(\hat{\boldsymbol{\theta}})\boldsymbol{U}_{i1}(\hat{\boldsymbol{\theta}})^\top\right\},$$

and the covariance matrix associated with estimating equation $\boldsymbol{U}_2(\boldsymbol{\theta}_0)$ can be consistently estimated by

$$\hat{\sigma}_{22}^2=\frac{\rho}{n}\sum_{i=1}^n\frac{\delta_i\sum_{j\in R_1(Y_i)}^n\left[\{Z_j,e(\mathbf{X}_j;\hat{\boldsymbol{\gamma}})\}^\top-\boldsymbol{\mu}_i\right]\exp\left\{\alpha Z_j+\hat{\beta}e(\mathbf{X}_j;\hat{\boldsymbol{\gamma}})\right\}}{\sum_{j\in R_1(Y_i)}^n\exp\left\{\alpha Z_j+\hat{\beta}e(\mathbf{X}_j;\hat{\boldsymbol{\gamma}})\right\}}$$

$$+\frac{(1-\rho)}{m}\sum_{i=1}^m\frac{\hat{\delta}_i\sum_{j\in R_2(\tilde{Y}_i)}^n\left[\{\tilde{Z}_j,e(\tilde{\mathbf{X}}_j;\hat{\boldsymbol{\gamma}})\}^\top-\tilde{\boldsymbol{\mu}}_i\right]\exp\left\{\alpha\tilde{Z}_j+\hat{\beta}e(\tilde{\mathbf{X}}_j;\hat{\boldsymbol{\gamma}})\right\}}{\sum_{j\in R_2(\tilde{Y}_i)}^n\exp\left\{\alpha\tilde{Z}_j+\hat{\beta}e(\tilde{\mathbf{X}}_j;\hat{\boldsymbol{\gamma}})\right\}},$$

where

$$\boldsymbol{\mu}_i = \frac{\sum_{k \in R_1(Y_i)} \{Z_k, e(\mathbf{X}_k; \hat{\boldsymbol{\gamma}})\}^{\top} \exp\{\alpha Z_k + \hat{\beta} e(\mathbf{X}_k; \hat{\boldsymbol{\gamma}})\}}{\sum_{k \in R_1(Y_i)} \exp\{\alpha Z_k + \hat{\beta} e(\mathbf{X}_k; \hat{\boldsymbol{\gamma}})\}},$$

$$\tilde{\boldsymbol{\mu}}_i = \frac{\sum_{k \in R_2(\tilde{Y}_i)} \{\tilde{Z}_k, e(\tilde{\mathbf{X}}_k; \hat{\boldsymbol{\gamma}})\}^{\top} \exp\{\alpha \tilde{Z}_k + \hat{\beta} e(\tilde{\mathbf{X}}_k; \hat{\boldsymbol{\gamma}})\}}{\sum_{k \in R_2(\tilde{Y}_i)} \exp\{\alpha \tilde{Z}_k + \hat{\beta} e(\tilde{\mathbf{X}}_k; \hat{\boldsymbol{\gamma}})\}}.$$

Summarizing equations (9) and (10), we have the asymptotic normality for $\hat{\boldsymbol{\theta}}$

$$\sqrt{N} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right) \to N(0, \boldsymbol{A}^{-1} \sum \boldsymbol{A}^{-1}).$$

A.3. Proof of Theorem 2

Without loss of generality, we assume $\mu\{e(\tilde{\mathbf{X}}_i; \hat{\boldsymbol{\gamma}})\} = e(\tilde{\mathbf{X}}_i; \hat{\boldsymbol{\gamma}})$ in the proof of Theorem 2. Let $\boldsymbol{\eta} = (\boldsymbol{\xi}, \boldsymbol{\zeta})$ with true value $\boldsymbol{\eta}_0$, and

$$\boldsymbol{U}_{\text{test}}(\boldsymbol{\eta}) = \sum_{i=1}^{n} \frac{\tilde{\delta}_i \left\{ e(\tilde{\mathbf{X}}_i; \hat{\boldsymbol{\gamma}}), \tilde{Z}_i \right\}^{\top}}{\hat{w}(\tilde{Y}_i)} \{\tilde{\mathbf{X}} - \boldsymbol{\xi} e(\tilde{\mathbf{X}}_i; \hat{\boldsymbol{\gamma}}) - \boldsymbol{\zeta} \tilde{Z}_i\}.$$

Applying the Taylor expansion of $\boldsymbol{U}_{test}(\boldsymbol{\eta})$ around $\boldsymbol{\eta}_0$:

$$\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) = \left\{ -\frac{1}{n} \frac{\partial}{\partial \boldsymbol{\eta}} \boldsymbol{U}_{\text{test}}(\boldsymbol{\eta}_0) \right\}^{-1} \frac{1}{\sqrt{n}} \boldsymbol{U}_{\text{test}}(\boldsymbol{\eta}_0) + o(1).$$

By the law of large numbers, $\frac{\partial}{\partial \boldsymbol{\eta}} \boldsymbol{U}_{\text{test}}(\boldsymbol{\eta}_0)/n$ converges in probability to its expectation, denoted as $\boldsymbol{A}_{\eta}$. The remaining task is to check the asymptotic behavior of $\boldsymbol{U}_{test}(\boldsymbol{\eta}_0)$. It can be shown that

$$\boldsymbol{U}_{\text{test}}(\boldsymbol{\eta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\tilde{\delta}_i \left\{ e(\hat{\mathbf{X}}_i; \hat{\boldsymbol{\gamma}}), \tilde{Z}_i \right\}^{\top}}{\hat{w}(\tilde{Y}_i)} \{\tilde{\mathbf{X}}_i - \boldsymbol{\xi} e(\tilde{\mathbf{X}}_i; \hat{\boldsymbol{\gamma}}) - \boldsymbol{\zeta} \tilde{Z}_i\}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\tilde{\delta}_i \left\{ e(\hat{\mathbf{X}}_i; \boldsymbol{\gamma}), \tilde{Z}_i \right\}^{\top}}{w(\tilde{Y}_i)} \{\tilde{\mathbf{X}}_i - \boldsymbol{\xi} e(\tilde{\mathbf{X}}_i; \boldsymbol{\gamma}) - \boldsymbol{\zeta} \tilde{Z}_i\} \tag{11}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\tilde{\delta}_i}{\hat{w}(\tilde{Y}_i)}-\frac{\tilde{\delta}_i}{w(\tilde{Y}_i)}\right\}\{e(\tilde{X}_i;\boldsymbol{\gamma}),\tilde{Z}_i\}^\top\{\tilde{\mathbf{X}}_i-\boldsymbol{\xi}e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma})-\boldsymbol{\zeta}\tilde{Z}_i\} \tag{12}$$

$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\tilde{\delta}_i\{e(\hat{\mathbf{X}}_i;\boldsymbol{\gamma}),\tilde{Z}_i\}^\top}{w(\tilde{Y}_i)}\{\boldsymbol{\xi}e(\tilde{\mathbf{X}}_i;\hat{\boldsymbol{\gamma}})-\boldsymbol{\xi}e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma})\} \tag{13}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\tilde{\delta}_i}{\hat{w}(\tilde{Y}_i)}-\frac{\tilde{\delta}_i}{w(\tilde{Y}_i)}\right\}\{e(\tilde{X}_i;\boldsymbol{\gamma}),\tilde{Z}_i\}^\top\{\boldsymbol{\xi}e(\tilde{\mathbf{X}}_i;\hat{\boldsymbol{\gamma}})-\boldsymbol{\xi}e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma})\} \tag{14}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\tilde{\delta}_i(e(\hat{\mathbf{X}}_i;\hat{\boldsymbol{\gamma}})-e(\hat{\mathbf{X}}_i;\boldsymbol{\gamma}))}{w(\tilde{Y}_i)}\{\tilde{\mathbf{X}}_i-\boldsymbol{\xi}e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma})-\boldsymbol{\zeta}\tilde{Z}_i\}. \tag{15}$$

By the uniform consistency of $\hat{w}(y)$ to $w(y)$ and the martingale expression of the Kaplan-Meier estimator, (12) can be rewritten as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{w(\tilde{Y}_i)-\hat{w}(\tilde{Y}_i)}{w^2(\tilde{Y}_i)}\tilde{\delta}_i\{e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma}),\tilde{Z}_i\}^\top\left\{\tilde{\mathbf{X}}_i-\boldsymbol{\xi}e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma})-\boldsymbol{\zeta}\tilde{Z}_i\right\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\int_0^{t_0}\boldsymbol{D}(t)dM_i(t)}{S_c(t)S_c(t)}\right\}+o(1),$$

where

$$\boldsymbol{D}(t)=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}I(t\le Y_i)\delta_i\{e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma}),Z_i\}^\top\int_t^{Y_i}\hat{S}_c(u)d(u)\frac{\{\tilde{\mathbf{X}}_i-\boldsymbol{\xi}e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma})-\boldsymbol{\zeta}\tilde{Z}_i\}}{\hat{w}^2(Y_i)},$$
$$M_i(t)=I(Y_i-W_i\le t,\delta_i=0)-\int_0^t I(Y_i-W_i\ge u)d\Lambda_c(u),$$

and $\Lambda_c(u)$ is the cumulative hazard function of the censoring time. By the martingale central limit theorem and regularity condition C.7), (12) follows a normal distribution with variance denoted as Σ_M. The asymptotic behavior of (13) is determined by the asymptotic behavior of $\hat{\boldsymbol{\gamma}}$, because it is equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\tilde{\delta}_i(1,\tilde{Z}_i)^\top}{w(\tilde{Y}_i)}\boldsymbol{\xi}\left\{\frac{\partial}{\partial\boldsymbol{\gamma}}e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma})\right\}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})+o(1). \tag{16}$$

By Theorem 1, when $m \to \infty$ and $n \to \infty$, we have $\sqrt{n}(\hat{\gamma} - \gamma) \to^d N(0, \sigma_\gamma^2)$, where $\sigma_\gamma^2 = (1 - \rho)\sigma_\gamma^*$ and σ_γ^* is the $(q + 1) \times (q + 1)$ upper submatrix of $A^{-1}\Sigma A^{-1}$. Therefore, (13) converges in distribution to a normal distribution with mean 0 and variance Σ_2, where

$$\sum\nolimits_2 = E\left\{ \frac{\tilde{\delta}_i(e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma}), \tilde{Z}_i)^\top}{w(\tilde{Y}_i)} \boldsymbol{\xi} \frac{\partial}{\partial \boldsymbol{\gamma}} e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma}) \right\}^\top \sigma_\gamma^2 E\left\{ \frac{\tilde{\delta}_i(e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma}), \tilde{Z}_i)^\top}{w(\tilde{Y}_i)} \boldsymbol{\xi} \frac{\partial}{\partial \boldsymbol{\gamma}} e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma}) \right\}.$$

After some algebra, we can show that (14) converges to 0 in probability. Similar to the arguments used for (13), (15) converges in distribution to a normal distribution.

Summarizing the previous arguments, $\sqrt{n}\{\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\} \to^d N(0, A_\eta^{-1}\sum_\eta A_\eta^{-1})$, where

$$\sum\nolimits_\eta = E\left[\frac{\tilde{\delta}_i(e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma}), \tilde{Z}_i)^\top}{w(\tilde{Y}_i)} \{\tilde{\mathbf{X}}_i - \boldsymbol{\xi} e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma}) - \boldsymbol{\zeta}\tilde{Z}_i\} + A_3 \mathbf{U}_{i1}(\boldsymbol{\gamma}) \right]^{\otimes 2} + \sum\nolimits_M$$

and

$$A_3 = E\left[\frac{\tilde{\delta}_i(e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma}), \tilde{Z}_i)^\top}{w(\tilde{Y}_i)} \boldsymbol{\xi} \frac{\partial}{\partial \boldsymbol{\gamma}} e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma}) + \frac{\tilde{\delta}_i}{w(\tilde{Y}_i)} \{\tilde{\mathbf{X}}_i - \boldsymbol{\xi} e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma}) - \boldsymbol{\zeta}\tilde{Z}_i\} \frac{\partial}{\partial \boldsymbol{\gamma}} e(\tilde{\mathbf{X}}_i;\boldsymbol{\gamma}) \right].$$

This implies that under the null hypothesis, $\hat{\boldsymbol{\zeta}}$ converges in distribution to a normal distribution with mean 0 and variance V, where V is the submatrix of $A_\eta^{-1}\sum_\eta A_\eta^{-1}$ corresponding to $\hat{\boldsymbol{\zeta}}$. In other words, the limit distribution of the proposed test T_{test} is a chi-squared limiting distribution with a degree of freedom q under the null hypothesis.

Bibliography

Asgharian M, Wolfson DB, et al. Asymptotic behavior of the unconditional npmle of the length-biased survivor function from right censored prevalent cohort data. The Annals of Statistics. 2005; 33:2109–2131.

Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. Statistics in medicine. 2008; 27:2037–2049. [PubMed: 18038446]

Cardoso F, Harbeck N, Fallowfield L, Kyriakides S, Senkus E, Group EGW, et al. Locally recurrent or metastatic breast cancer: Esmo clinical practice guidelines for diagnosis, treatment and follow-up. Annals of oncology. 2012; 23:vii11–vii19. [PubMed: 22997442]

Chan KCG. Survival analysis without survival data: connecting length-biased and case-control data. Biometrika. 2013; 100:1–7.

Cheng YJ, Wang MC. Estimating propensity scores and causal survival functions using prevalent survival data. Biometrics. 2012; 68:707–716. [PubMed: 22834993]

D'Agostino RB. Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Statistics in medicine. 1998; 17:2265–2281. [PubMed: 9802183]

Ertefaie A, Asgharian M, Stephens D. Propensity score estimation in the presence of lengthbiased sampling: a non-parametric adjustment approach. Stat. 2014; 3(1):83–94. [PubMed: 25170178]

Gail, MH., Benichou, J. Encyclopedia of epidemiologic methods. John Wiley & Sons; 2000.

Gonzaález-Manteiga W, Cadarso-Suarez C. Asymptotic properties of a generalized kaplan-meier estimator with some applications. Communications in Statistics-Theory and Methods. 1994; 4:65–78.

Hansen BB, Bowers J. Covariate balance in simple, stratified and clustered comparative studies. Statistical Science. 2008; 23:219–236.

Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. Journal of the royal statistical society: series A (statistics in society). 2008; 171:481–502.

Kang JD, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. Statistical science. 2007; 22:523–539.

Lai TL, Ying Z. Rank regression methods for left-truncated and right-censored data. The Annals of Statistics. 1991; 10:531–556.

Li L, Greene T. A weighting analogue to pair matching in propensity score analysis. The international journal of biostatistics. 2013; 9:215–234. [PubMed: 23902694]

Ning J, Qin J, Shen Y. Buckley–james-type estimator with right-censored and length-biased data. Biometrics. 2011; 67:1369–1378. [PubMed: 21385160]

Qin J, Shen Y. Statistical methods for analyzing right-censored length-biased data under cox model. Biometrics. 2010; 66:382–392. [PubMed: 19522872]

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70:41–55.

Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician. 1985; 39:33–38.

Rothman, KJ., Greenland, S., Lash, TL. Modern epidemiology. Lippincott Williams & Wilkins; 2008.

Rubin DB. Matching to remove bias in observational studies. Biometrics. 1973; 29:159–183.

Rubin DB. Using multivariate matched sampling and regression adjustment to control bias in observational studies. Journal of the American Statistical Association. 1979; 74:318–328.

Shen Y, Ning J, Qin J. Analyzing length-biased data with semiparametric transformation and accelerated failure time models. Journal of the American Statistical Association. 2009; 104:1192–1202. [PubMed: 21057599]

Tsiatis AA. A large sample study of cox's regression model. The Annals of Statistics. 1981; 9:93–108.

Vardi Y. Nonparametric estimation in the presence of length bias. The Annals of Statistics. 1982; 10:616–620.

Wang MC, Brookmeyer R, Jewell NP. Statistical models for prevalent cohort data. Biometrics. 1993; 49:1–11. [PubMed: 8513095]

Zelen, M. Probability, Statistics and Modelling in Public Health. Springer; 2006. Forward and backward recurrence times and length biased sampling: age specific models; p. 1-11.
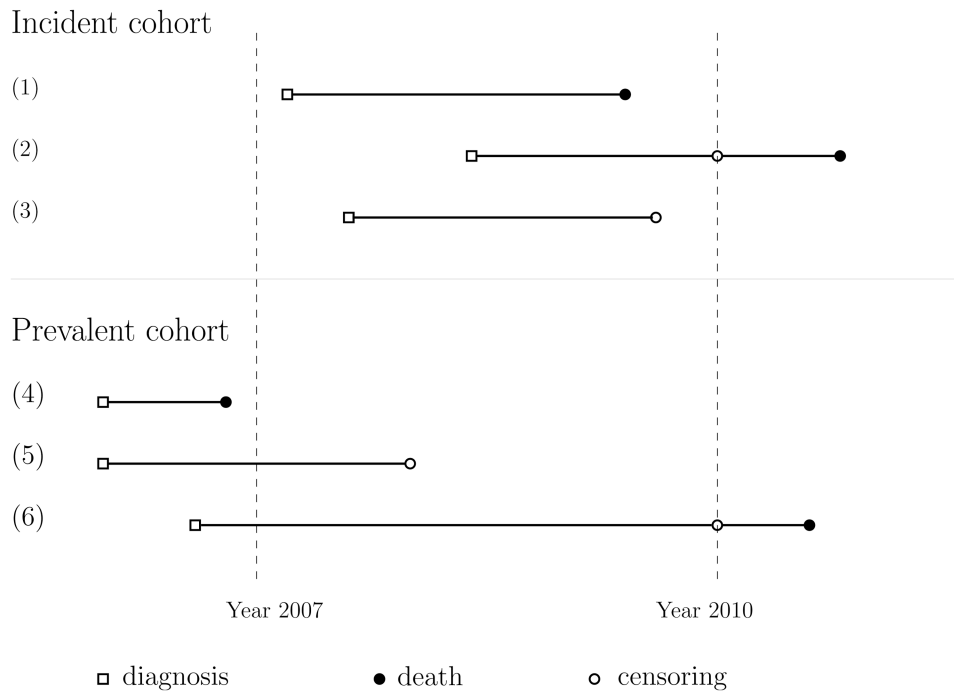
Figure 1. Incident and prevalent cohort sampling schemas

Table 1

Summary of 1, 000 simulations of the proposed method by combining incident and prevalent cohorts: empirical bias (Bias), empirical standard error (ESE), asymptotic standard error (ASE), and coverage probability (CP).

Sample Size (m,n)		Scenario I: $C_1\% = 60\%$, $C_2\% = 40\%$						Scenario II: $C_1\% = 40\%$, $C_2\% = 20\%$					
		$\gamma = (0.1, 0.4, 0.4, 0.4)$				$(\alpha, \beta) = (0.5,1)$		$\gamma = (0.1, 0.4, 0.4, 0.4)$				$(\alpha, \beta) = (0.5,1)$	
(200, 200)	Bias	-.003	.021	.008	.018	.028	-.007	-.003	.021	.008	.018	.023	-.021
	ESE	.163	.157	.156	.155	.155	.576	.163	.157	.156	.155	.127	.482
	ASE	.152	.158	.159	.159	.160	.583	.152	.158	.159	.159	.134	.498
	CP	.938	.958	.956	.970	.964	.962	.938	.958	.956	.970	.964	.962
(400, 200)	Bias	.000	.014	-.003	-.005	.012	-.013	.000	.014	-.003	-.005	.012	.006
	ESE	.102	.114	.118	.114	.136	.463	.102	.114	.118	.114	.112	.395
	ASE	.106	.110	.111	.110	.133	.460	.106	.110	.111	.110	.111	.392
	CP	.960	.946	.912	.944	.940	.960	.960	.946	.912	.944	.952	.952
(600, 200)	Bias	.002	.010	.008	.007	.018	.002	.002	.010	.008	.007	.015	-.004
	ESE	.086	.090	.090	.086	.125	.402	.086	.090	.090	.086	.105	.338
	ASE	.087	.090	.090	.090	.117	.389	.087	.090	.090	.090	.097	.328
	CP	.948	.954	.948	.958	.932	.946	.948	.954	.948	.958	.930	.948

$C_1\%$ and $C_2\%$: the respective censoring rates of the incident cohort and prevalent cohort

Table 2

Summary of 1, 000 simulations of naive methods: empirical bias (Bias), empirical standard error (ESE), asymptotic standard error (ASE), and coverage probability (CP).

Sample Size		Scenario I: $C_1\% = 60\%$, $C_2\% = 40\%$						Scenario II: $C_1\% = 40\%$, $C_2\% = 20\%$					
		$\gamma = (0.1, 0.4, 0.4, 0.4)$				$(\alpha, \beta) = (0.5, 1)$		$\gamma = (0.1, 0.4, 0.4, 0.4)$				$(\alpha, \beta) = (0.5, 1)$	
Method using data from incident cohort only													
200	Bias	-.003	.021	.008	.018	.005	-.001	.018	.008	.021	-.003	.011	-.029
	ESE	.163	.157	.156	.155	.248	.876	.155	.156	.157	.163	.192	.704
	ASE	.152	.158	.159	.159	.249	.867	.159	.159	.158	.152	.203	.720
	CP	.938	.958	.956	.970	.958	.970	.970	.956	.958	.938	.954	.966
400	Bias	.000	.014	-.003	-.005	.006	-.024	-.005	-.003	.014	.000	.009	-.007
	ESE	.102	.114	.118	.114	.170	.568	.114	.118	.114	.102	.142	.481
	ASE	.106	.110	.111	.110	.174	.583	.110	.111	.110	.106	.141	.487
	CP	.960	.946	.912	.944	.952	.974	.944	.912	.946	.960	.956	.960
600	Bias	.002	.010	.008	.007	.014	-.015	.007	.008	.010	.002	.010	-.009
	ESE	.086	.090	.090	.086	.149	.478	.086	.090	.090	.086	.119	.381
	ASE	.087	.090	.090	.090	.142	.463	.090	.090	.090	.087	.115	.384
	CP	.948	.954	.948	.958	.938	.938	.958	.948	.954	.948	.940	.948
Method using data from prevalent cohort only													
200	Bias	-.503	.011	.012	.019	.315	.638	.019	.012	.011	-.503	.262	.555
	ESE	.170	.166	.165	.174	.205	.783	.174	.165	.166	.170	.177	.683
	ASE	.158	.162	.163	.163	.209	.836	.163	.163	.162	.158	.180	.731
	CP	.120	.944	.944	.940	.664	.950	.940	.944	.944	.120	.694	.950
400	Bias	-.506	.007	.006	.006	.303	.581	.006	.007	.006	-.506	.258	.505
	ESE	.111	.126	.108	.111	.140	.559	.111	.108	.126	.111	.124	.469
	ASE	.110	.114	.114	.114	.146	.548	.114	.114	.114	.110	.125	.479
	CP	.002	.916	.958	.968	.470	.866	.968	.958	.916	.002	.472	.904
600	Bias	-.495	.001	.001	.002	.292	.612	.002	.001	.001	-.495	.251	.516
	ESE	.084	.092	.099	.095	.119	.422	.095	.099	.092	.084	.104	.370
	ASE	.090	.092	.092	.092	.118	.442	.092	.092	.092	.090	.102	.387
	CP	.000	.950	.944	.938	.298	.764	.938	.944	.950	.000	.320	.786

Author Manuscript

Author Manuscript

C_1% and C_2% respectively denote the censoring rate of the incident cohort and prevalent cohort

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Simulation results: rates of rejection of the null hypothesis

(m, n)	$C_1\% = 60\%, C_2\% = 40\%$			$C_1\% = 40\%, C_2\% = 20\%$		
	Type I	Power$_1$	Power$_2$	Type I	Power$_1$	Power$_2$
(200,200)	0.070	0.672	0.213	0.063	0.704	0.256
(400,200)	0.043	0.683	0.173	0.058	0.711	0.220
(400,400)	0.049	0.873	0.223	0.058	0.877	0.272
(600,600)	0.046	0.945	0.240	0.053	0.949	0.296

Table 4

Distribution of patient characteristics by cohort

Variable	Incident ($m = 1867$)	Prevalent ($n = 1106$)	*P*-value
Median survival time (years)	3.6	7.6	
Age at diagnosis (years)	77.2	75.9	< 0.0001
Race			
Non-Hispanic white	1494 (80.0%)	893 (80.7%)	0.6678
Non-Hispanic black	206 (11.0%)	117 (10.6%)	0.7149
Hispanic and other	167 (8.9%)	96 (8.7%)	0.8413
Comorbidity			
0	1207 (64.6%)	808 (73.1%)	< 0.0001
1	406 (21.7%)	185 (16.7%)	< 0.0008
2	254 (13.6%)	113 (10.2%)	0.0068
ER			
Positive	1136 (60.8%)	784 (70.9%)	< 0.0001
Negative	349 (18.7%)	152 (13.7%)	0.0005
Other	382 (20.5%)	170 (15.4%)	0.0005
PR			
Positive	871 (46.7%)	606 (54.8%)	< 0.0001
Negative	588 (31.5%)	307 (27.8%)	0.0349
Other	408 (21.9%)	193 (17.5%)	0.0039

Table 5

Estimation results by Cox models using the cohorts from the SEER-Medicare linked data. HR: hazard ratio; SE: standard error

	Incident cohort			Combined cohort		
	HR	**SE**	**P-value**	**HR**	**SE**	**P-value**
Radiation effect without adjusting by propensity score [*]	0.657	0.033	< 0.001	0.700	0.028	< 0.001
Radiation effect adjusted by propensity score	0.756	0.045	< 0.001	0.748	0.030	< 0.001

[*]
the model does not include covariates X.