Published in final edited form as:

J Appl Stat. 2012 ; 39(4): 901–911. doi:10.1080/02664763.2011.626849.

Change point-cure models with application to estimating the change-point effect of age of diagnosis among prostate cancer patients

Megan Othus^{a,*}, Yi Li^b, and Ram Tiwari^c

^a1100 Fairview Ave N, M3-C102, PO Box 19024, Seattle WA 98117-1024, USA

^bDepartment of Biostatistics, Harvard University and Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston MA, USA

^cOffice of Biostatistics, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Springs MD, USA

Abstract

Previous research on prostate cancer survival trends in the United States National Cancer Institute's Surveillance Epidemiology and End Results database has indicated a potential change-point in the age of diagnosis of prostate cancer around age 50. Identifying a change-point value in prostate cancer survival and cure could have important policy and health care management implications. Statistical analysis of this data has to address two complicating features: (1) change-point models are not smooth functions and so present computational and theoretical difficulties; and (2) models for prostate cancer survival need to account for the fact that many men diagnosed with prostate cancer can be effectively cured of their disease with early treatment. We develop a cure survival model that allows for change-point effects in covariates to investigate a potential change-point in the age of diagnosis of prostate cancer. Our results do not indicate that age under 50 is associated with increased hazard of death from prostate cancer.

Keywords

cure models; change-point models; right censoring; smoothing; survival analysis

1. Introduction

Previous research has indicated that younger age is a prognostic factor for survival after diagnosis with prostate cancer, with age under 50 at diagnosis associated with poor survival [24, hereinafter denoted MB]. The trends were strongest for men with distant disease, but were also present for men with local/regional disease. Given that over 75% of men in the United States National Cancer Institute's Surveillance Epidemiology and End Results (SEER) database were diagnosed with local or regional disease, we are interested in whether an analysis that explicitly allows for a change-point could elucidate further information about the relationship between age at diagnosis and prostate cancer survival. Specifically, we are interested in whether there is statistical evidence in the SEER prostate cancer data for a change-point in the age at diagnosis for men diagnosed with local or regional disease.

Identifying when and if a change-point exists could improve treatment planning and help to direct resource allocation.

Statistical analysis of this data involves accounting for two complicating features: (1) many men diagnosed with early stage prostate cancer can be cured of their disease with chemotherapy or radiation treatment, and (2) change-points are by definition not smooth functions and so present computational and theoretical challenges. First we outline relevant prior research.

Progress in cancer diagnosis and treatment has led to a considerable amount of statistical research for survival data with cure fractions. A popular approach for this type of data is to assume the population is a mixture of cured and not cured subjects. Related work includes parametric mixture models [3, 6], a mixture of the proportional hazards model and logistic regression [17, 27, 30], semiparametric transformation models combined with logistic regression [20], and copula models that allow for dependent censoring [18]. Bayesian inference has focused on the promotion cure model [5, 31], which has an intuitive biological motivation. All of the models referenced above assume that the covariates act smoothly on the cure rate or the hazard rate.

Unfortunately, these models are not appropriate when covariates have change-point effects on the cure or hazard rates. Models that accommodate a change-point in the hazard function have been considered by a number of authors [7, 8, 10, 13, 15, 19, 21–23, 25, 29, 34]. However, these models do not accommodate populations that have a cure fraction.

There does not exist a model that addresses issues (1) and (2) outlined above, so along with analyzing the SEER prostate cancer data, this project develops a novel cure survival model that allows for a change-point effect in covariates. When we apply the proposed model to the SEER data, we do not find evidence that younger age is a prognostic factor for men with local/regional disease. We estimate the change-point for both black and white men to be over 70 years old. Older age is associated with poorer survival for both races.

The rest of the article is organized as follows: Section 2 introduces notation and the new model; the smoothed approach to the problem is explained and asymptotic results are summarized in Section 3; Section 4 discusses computational issues and provides a summary of simulation studies; the analysis of the SEER prostate cancer data is presented in Section 5; the paper ends with a brief summary in Section 6. Variance formulae are provided in the Appendix. A full theoretical development of the proposed model is provided in a Technical Report available from the first author.

2. Model Specification

Denote the survival time for a subject by T . In this project we take T to be the time from diagnosis of prostate cancer to death from prostate cancer, though the method is applicable to general survival times. The random variable T is assumed to have the following decomposition: $T = \nu T^* + (1 - \nu)\infty$, where $\nu \in \{0, 1\}$ is an unobservable indicator of whether the sampled subject is cured ($\nu = 0$) or not ($\nu = 1$), and T^* , with support on $[0, \infty)$, denotes the failure time if the subject is not cured. Let C denote the censoring time and let X be a covariate related to T^* and the cure indicator ν . Define $\tilde{T} = \min(T, C)$ and $\Delta = I(T < C)$, where $I(\cdot)$ is the indicator function, with $I(A)$ taking the value 1 when the event A is true and the value 0 otherwise. In practice, we observe $(\tilde{T}_i, \Delta_i, X_i)$ for $i = 1, \dots, n$, which are independent and identically distributed copies of $(\tilde{T}, \Delta, X)$. We assume that T and C are independent, conditional on X . For now we assume that X is a scalar, but details for extensions to multiple covariates are available in the Technical Report.

Define the hazard for patients who are not cured as λ and the probability a patient is cured as η . We assume that the change-point of the model depends on X , and that at the change-point the hazard value or cure probability takes a sudden jump or fall. Suppose that the change-point for X is τ . If $X \leq \tau$, write $\eta(X) = c_1$ and $\lambda(X) = \lambda_1$, while if $X > \tau$, let $\eta(X) = c_2$ and $\lambda(X) = \lambda_2$ for some constants $|c_1 - c_2| + |\lambda_1 - \lambda_2| > 0$. In other words, let

$$\eta(X) = c_1 I(X \leq \tau) + c_2 I(X > \tau) \text{ and } \lambda(X) = \lambda_1 I(X \leq \tau) + \lambda_2 I(X > \tau).$$

This model is an extension of the change-point hazard rate model proposed by Matthews and Farewell [23] to accommodate a cure fraction. The survival function for T , known as a cure model or cure survival function, can be written

$$S^*(t|X) = P(T \geq t|X) = \{1 - \eta(X)\} \exp\{-t\lambda(X)\} + \eta(X). \quad (1)$$

The survival function for those not cured is often referred to as the latency survival function. Based on Equation (1), the density of T at t can be written $\lambda(X)\{1 - \eta(X)\} \exp\{-t\lambda(X)\}$.

Define the unknown parameters by $\theta = (\varphi, \tau) = (c_1, c_2, \lambda_1, \lambda_2, \tau)$. The likelihood associated with the observed data, denoted by $L_n^*(\theta)$, is

$$\begin{aligned} & \prod_{i=1}^n \left[\lambda(X_i) \{1 - \eta(X_i)\} \exp\{-\tilde{T}_i \lambda(X_i)\} \right]^{\Delta_i} \left[\{1 - \eta(X_i)\} \exp\{-\tilde{T}_i \lambda(X_i)\} + \eta(X_i) \right]^{1-\Delta_i} \\ &= \prod_{i=1}^n \left[\{\lambda_1 (1 - c_1) \exp(-\tilde{T}_i \lambda_1)\}^{\Delta_i} \{(1 - c_1) \exp(-\tilde{T}_i \lambda_1) + c_1\}^{1-\Delta_i} \right]^{I(X_i \leq \tau)} \\ & \quad \times \left[\{\lambda_2 (1 - c_2) \exp(-\tilde{T}_i \lambda_2)\}^{\Delta_i} \{(1 - c_2) \exp(-\tilde{T}_i \lambda_2) + c_2\}^{1-\Delta_i} \right]^{I(X_i > \tau)}. \end{aligned} \quad (2)$$

Because τ is unknown and the likelihood is not differentiable with respect to τ , standard Taylor series methods cannot be used to develop asymptotic theory for $L_n^*(\theta)$ and consistency results are difficult to obtain. Also, the convergence rate of the change-point is difficult to assess from the likelihood.

3. Smoothed Likelihood Approach

To address analytic issues associated with $L_n^*(\theta)$, we propose a smoothed likelihood approach to the problem. The key idea is to use a continuous differentiable function to approximate the indicator functions $I(X \leq \tau)$ and $I(X > \tau)$. Let $K(u)$ be a continuous function such that $K(u)$ is differentiable and non-decreasing over the real line, with $\lim_{u \rightarrow -\infty} K(u) = 0$ and $\lim_{u \rightarrow \infty} K(u) = 1$. Define $K_n(u) = K(u/h_n)$, where h_n is a small real number that depends on the sample size. Section 4.1 will discuss how to select h_n . For sufficiently small h_n , $K_n(\tau - X)$ approaches 1 when $\tau - X > 0$ and is close to 0 when $\tau - X < 0$, so the change-point feature is preserved in the smoothed model. For the remainder of this paper, we shall consider a useful and often used special case of this class of functions, the logistic function, where $K(u) = \exp(u) / \{1 + \exp(u)\}$.

The smoothed likelihood associated with the observed data, denoted by $L_n(\theta)$, is

$$\prod_{i=1}^n [\{\lambda_1(1-c_1)\exp(-\tilde{T}_i\lambda_1)\}^{\Delta_i} \{(1-c_1)\exp(-\tilde{T}_i\lambda_1)+c_1\}^{1-\Delta_i}]^{K_n(\tau-X_i)} \times [\{\lambda_2(1-c_2)\exp(-\tilde{T}_i\lambda_2)\}^{\Delta_i} \{(1-c_2)\exp(-\tilde{T}_i\lambda_2)+c_2\}^{1-\Delta_i}]^{1-K_n(\tau-X_i)} \quad (3)$$

and the maximizer of $L_n(\theta)$ is denoted by $\hat{\theta}$. In contrast to Equation (2), $L_n(\theta)$ is differentiable with respect to θ .

Denote the true parameter values of θ with θ_0 , the log-likelihood associated with Equation (3) with $l_n(\theta)$, the Euclidean norm with $|\cdot|$, first derivatives with superscript dots, and second derivatives with double superscript dots. The following regularity conditions are used in the asymptotic results:

- C.1** the parameter space, Θ , containing θ_0 is a compact set of $\mathbb{R}^5$;
- C.2** $c_1, c_2 \in (0, 1)$;
- C.3** $|\lambda_1 - \lambda_2| + |c_1 - c_2| > 0$ and a change-point actually occurs at τ_0 ;
- C.4** X is continuous and uniformly bounded, the density of X , $f_X(x)$, is every-where positive, and $\dot{f}_X(x)$ and $\ddot{f}_X(x)$ are Lipschitz;
- C.5** $E\{l_n(\theta)\}$ is a continuous function of θ ;
- C.6** the density for T is continuous almost everywhere and its density has continuous first and second derivatives;
- C.7** ${}^2E\{l_n(\theta)\}/\theta^2$ is nondegenerate;
- C.8** τ_0 is in a bounded interval strictly included in the support of X .

Conditions C.1, C.5, C.6, and C.7 are standard assumptions in survival analysis. Condition C.2 assumes that the population is a mixture of cured and not cured subjects and is a common assumption in cure modeling. Condition C.3 is needed for identifiability. Condition C.4 is needed to prove consistency. Condition C.8 ensures that τ can be estimated.

Under regularity conditions C.1–C.8, we have the following properties for the proposed model and estimates. Proofs are presented in the Technical Report.

Theorem 3.1

Under Conditions C.1 – C.4, the survival model (1) is identifiable.

Theorem 3.2

Under Conditions C.1 – C.7, if $h_n \rightarrow 0$ as $n \rightarrow \infty$, $\hat{\theta} \rightarrow \theta_0$ almost surely.

Theorem 3.3

Under Conditions C.1 – C.7, if $h_n^2(nh_n)^{1/2} \rightarrow 0$

$$n^{1/2}(\hat{\varphi} - \varphi_0) \rightarrow N(0, \tilde{\Sigma}_{\varphi}), \quad (nh_n)^{1/2}(\hat{\tau} - \tau_0) \rightarrow N(0, \tilde{\Sigma}_{\tau})$$

in distribution for $n \rightarrow \infty$, where $\tilde{\Sigma}_{\varphi}$ and $\tilde{\Sigma}_{\tau}$ are defined in the Appendix.

Theorems 3.2 and 3.3 are important as they demonstrate that even though we have used a smoothed approximation for the change-point indicator, $\hat{\phi}$ has a centered distribution and converges at the usual $n^{1/2}$ rate. The convergence rate of the change-point estimate depends on the smoothing parameter h_n , as is usual for smoothing situations. Also of interest is the fact that $\hat{\phi}$ and $\hat{\tau}$ are asymptotically independent; a result that has precedence in other change-point problems [16, 29].

4. Computational Considerations

4.1 The Smoothing Parameter h_n

We propose to jointly estimate the parameters for the cure and latency survival functions, ϕ , and the change-point parameter, τ , by maximizing the log-likelihood associated with Equation (3), which can be accomplished using Newton-Raphson methods. A key component of the log-likelihood is the smoothing parameter h_n .

There is an inherent trade-off between bias and variance depending on the rate of convergence of the smoothing parameter h_n . As is the case with the general smoothing situation, smaller values for h_n lead to unbiased estimates but larger variances. Large values for h_n lead to a smaller relative variance, but also an increase in bias that leads to a non-centered asymptotic distribution. We will use $h_n = n^{-1/3}$, which we found to perform satisfactorily in simulations and which allows us to have a centered asymptotic distribution for τ . Cross-validation could be used to select h_n , but our simulations indicate that using the plug-in value works well and decreases computation time.

4.2 Algorithm

Step 1—Estimate initial values, $\theta^{(0)}$, for θ . A discussion for a choice of initial values can be found below.

Step 2—Solve for $\hat{\theta}$ as the maximum of $L_n(\theta)$ (Equation (3)). Standard optimization procedures, such as Newton-Raphson, can be used.

Step 3—Estimate variances or calculate confidence intervals. Variance formulae from the smoothed likelihood can be found in the Appendix. If small sample size is a concern, the BC_a bootstrap confidence interval may be used [4].

For the initial values for ϕ , we use naïve methods that ignore potential cure fractions and change-point effects in the data. For the initial estimates of the cure fractions, we use the tail of the Kaplan-Meier estimate of the survival function with $\hat{c}_1 = \hat{c}_2$. Similarly, for the initial estimates of λ_1 and λ_2 , we take $\hat{\lambda}_1 = \hat{\lambda}_2$ and estimate the survival parameter with the intercept coefficient of an exponential regression model with no covariates. A coarse grid search along the range of τ for the value that minimizes the log-likelihood can be used to find a reasonable initial estimate for τ .

4.3 Simulations

We conducted simulation studies to verify that our method performs adequately in finite samples. Two studies are summarized here. The first simulation uses a Uniform(0, 1) random variable with a change-point at 0.5. The second simulation uses a truncated Normal(1, 1, 0, 2) (tN(1, 1, 0, 2)) random variable with a change-point at 1. With this notation tN(0, 1, $-\infty$, ∞) denotes a standard normal random variable. Parameter values were chosen to mimic our data application after rescaling. Censoring times follow an Exponential(0.25) distribution, which induces a censoring rate of 45%.

Simulation results are based on 500 replications of sample sizes 250, 500, 1000, and 3000. Results are presented in Table A1. Monte Carlo (SE_{mc}) and analytic standard errors (SE) are reported along with the average of the estimates for each parameter (Est). The method has nearly unbiased results with reasonable standard errors. Even with a moderate sample size of 250, the estimate of the change-point is very accurate. Asymptotically the rate of convergence of τ is slower than for the other parameters. This most visibly manifests itself in the standard error values, where there are some discrepancies between the analytic and Monte Carlo standard errors for smaller sample sizes. That said, with a sample size of 500 the Monte Carlo SEs are close to the analytic SEs.

Horowitz [12] commented that large sample sizes were needed to observe the asymptotic properties of smoothed parameters. If a particular application has moderate sample sizes and an interval estimate for τ is desired, we recommend the use of the BC_a bootstrap interval; a description of this type of interval and justification for its use in this situation can be found in [4]. In simulation studies we found the BC_a intervals had nominal coverage for ϕ at sample sizes as small as 100, and achieved nominal coverage for τ with sample sizes around 500. Near nominal-coverage for τ was found for sample sizes as small as 250.

5. The SEER Prostate Cancer Data

The motivation for the proposed model was an application to SEER prostate cancer data. MB [24] found that patients diagnosed with prostate cancer before age 50 had poorer survival. The trends were strongest for men with distant disease, but were also present for men with local/regional disease. We are interested in whether an analysis that explicitly allows for a change-point could elucidate further information about the relationship between age at diagnosis and prostate cancer survival. Specifically, we are interested in whether there is statistical evidence in the SEER prostate cancer data for a change-point in the age at diagnosis (X in the following analysis) for men diagnosed with local or regional disease.

We restricted our analysis to men whose prostate cancer diagnosis was their first cancer diagnosis. To avoid potential biases associated with the implementation of the prostate-specific antigen test for screening for prostate cancer, we restricted our data to men diagnosed before 1990. We wished to look at different racial subgroups due to previously observed racial disparities in prostate cancer [1, 11, 26, 28, 32, 33, among others]. The issue of racial disparities is complicated and would be unable to be addressed given the limited nature of the SEER data, so we will perform a stratified analysis as an alternative. Due to small numbers for some groups we were only able to keep two racial categories, white and black. There are a total of 85,004 men in this cohort; 8,079 are classified as black and 76,925 are classified as white. The survival time is defined to be the time from diagnosis of prostate cancer to death due to prostate cancer, and is scaled to months/100 in our algorithm.

A Kaplan-Meier estimate for the survival function for this cohort is provided in Figure 1. Long follow-up, over 25 years, is easily seen. The survival curve plateaus at the end of the plot, giving evidence that a non-negligible portion of the population is not susceptible to dying of prostate cancer.

There are several situations when physicians consider patients to be cured of prostate cancer. First, surgery (radical prostatectomy) and radiotherapy (externalbeam radiotherapy, brachytherapy, or both) in early stage patients can entirely eradicate the disease, inducing a cure among a significant proportion of patients [9, 14]. Also, improvements in prostate cancer treatment have made it so that many prostate cancer patients who are not cured are treated for chronic disease rather than acute disease, and their probability of dying of prostate cancer is not elevated above that of the general population [2]. Given these

considerations, a model that explicitly accounts for a cure proportion and allows for inference on the cure proportion can be of use for this prostate cancer application.

Parameter estimates and 95% confidence intervals for the two racial groups for the proposed model are provided in Table A2.

For the black cohort, the estimate of the change-point age at diagnosis was 71. We estimate that 37% of men diagnosed with prostate cancer at or before age 71 are cured, while 32% of men diagnosed after age 71 are cured. Our model estimates that black men diagnosed with prostate cancer at or before age 71 who are not cured have survival following an Exponential(0.37) distribution with median survival of 15.6 years, while survival for those diagnosed after age 71 follows an Exponential(0.55) distribution with median survival of 10.5 years. The change-point value split this cohort almost in half: 4,382 men were diagnosed at or before age 71 and 3,697 men were diagnosed after age 71.

For the white cohort the change-point age at diagnosis was 72. There were 36,820 men diagnosed at or before age 72 and 40,105 men diagnosed after age 72. We estimate that 47% of white men diagnosed with prostate cancer at or before age 72 are cured, while 39% of white men diagnosed after age 72 are cured. Among the portion of this cohort who are not cured, men diagnosed with prostate cancer at or before age 72 have survival following an Exponential(0.44) distribution with median survival of 13.1 years, while men diagnosed after age 72 have survival following an Exponential(0.54) distribution with median survival of 10.7 years.

Plots of both the Kaplan-Meier survival function and the survival function based on our cure model are provided in Figure 2. We provide plots for both races broken down by their estimated change-point value. Our cure change-point model appears to fit the data reasonably well.

6. Discussion

Previous research has indicated that there may be a change-point in the age of diagnosis of prostate cancer, with men diagnosed at a younger age having worse survival than men diagnosed at an older age. However, the previous research did not account for cure fractions or model change-point features explicitly. Using our model, we found that older men had worse survival than younger men for both white and black races. The research in Merrill and Bird [24] suggested that the change-point effect may be for men diagnosed before age 50. In this application, using a cure survival model that allows for a change-point, estimates for the change-points in age of diagnosis were 71 for black men and 72 for white men, both much older than 50.

Our analysis estimated that under 50% of men diagnosed with local or regional prostate cancer are cured of their disease, regardless of race and age. This result suggests two items for future prostate cancer research. First, we should continue efforts at improving the proportion of men cured of prostate cancer. Second, a non-negligible portion of men diagnosed with local or regional prostate cancer are cured of their disease and regression models for prostate cancer survival should incorporate this feature of the data.

The cure change-point model we develop assumes that there exists a non-zero proportion of subjects who are cured. Before fitting such a model, it is important that there exists a rationale for the existence of a cure fraction from the underlying science of the problem. If estimates of the cure fraction are very close to zero, it may be evidence that a cure model is not appropriate for the data.

A formal test for the existence of a change-point is beyond the scope of this paper, but is an important area that needs to be explored. A number of authors have discussed change-point tests in the context of survival data [10, 16, 23, 25, 34], but no gold-standard test exists. Adaptations of existing tests performed poorly in simulation studies for our type of data.

While we believe the parametric model proposed fit the SEER data adequately, a more general model may be useful for datasets where the proposed model is too restrictive. Extensions to more complicated models for the cure and survival functions are a subject of future research. Instead of a constant cure proportion, a logistic formulation could be employed with a change-point in one covariate. A variety of models could be taken for the latency survival function. Semiparametric models could be a useful extension.

Acknowledgments

The authors thank two referees for comments that led to an improved paper. This work was supported in part by grant CA090998 from the National Cancer Institute. Most of R. Tiwari's work was done during his previous employment at the National Cancer Institute. The views expressed in this article do not necessarily represent those of the U.S.A. Food and Drug Administration or the National Cancer Institute.

References

1. Albain K, Unger J, Crowley J, Coltman C Jr, Hershman D. Racial disparities in cancer survival among randomized clinical trials patients of the Southwest Oncology Group. *Journal of the National Cancer Institute*. 2009; 101:984. [PubMed: 19584328]
2. American Cancer Society. Prostate cancer: Detailed guide. 2007. <http://www.cancer.org/Cancer/Prostate-Cancer/DetailedGuide/index>, last revised November, 2009
3. Berkson J, Gage R. Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*. 1952; 47:501–515.
4. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine*. 2000; 19:1141–64. [PubMed: 10797513]
5. Chen M, Ibrahim J, Sinha D. A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*. 1999; 94:909–910.
6. Farewell V. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*. 1982; 38:1041–1046. [PubMed: 7168793]
7. Ghosh J, Joshi S, Mukhopadhyay C. Asymptotics of a Bayesian approach to estimating change-point in a hazard rate. *Communications in Statistics-Theory and Methods*. 1996; 25:3147–3166.
8. Ghosh, J.; Joshi, S.; Mukhopadhyay, C. *Frontiers in Reliability: A Volume Commemorating the First 25 Years of the Indian Association for Productivity, Quality & Reliability*. World Scientific; 1998. Inference about sharp changes in hazard rates; p. 133-142.
9. Hanlon A, Hanks G. Failure patterns and hazard rates for failure suggest the cure of prostate cancer by external beam radiation. *Urology*. 2000; 55:725–729. [PubMed: 10792090]
10. Henderson R. A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika*. 1990; 77:835–843.
11. Hoffman R, Gilliland F, Eley J, Harlan L, Stephenson R, Stanford J, Albertson P, Hamilton A, Hunt W, Potosky A. Racial and ethnic differences in advanced-stage prostate cancer: the Prostate Cancer Outcomes Study. *Journal of the National Cancer Institute*. 2001; 93:388. [PubMed: 11238701]
12. Horowitz J. A smoothed maximum score estimator for the binary response model. *Econometrica*. 1992; 79:505–531.
13. James B, James K, Siegmund D. Tests for a change-point. *Biometrika*. 1987; 74:71–83.
14. Jani A, Hellman S. Early prostate cancer: clinical decision-making. *The Lancet*. 2003; 361:1045–1053.
15. Jensen U, Lutkebohmert C. A Cox-type regression model with change-points in the covariates. *Lifetime Data Analysis*. 2008; 14:276–285.

16. Kosorok M, Song R. Inference under right censoring for transformation models with a change-point based on a covariate threshold. *The Annals of Statistics*. 2007; 35:957–989.
17. Kuk A, Chen C. A mixture model combining logistic regression with proportional hazards regression. *Biometrika*. 1992; 79:531.
18. Li Y, Tiwari R, Guha S. Mixture Cure Survival Models with Dependent Censoring. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2007; 69:285–306.
19. Liang K, Self S, Liu X. The Cox proportional hazards model with change point: an epidemiologic application. *Biometrics*. 1990; 46:783–793. [PubMed: 2242414]
20. Lu W, Ying Z. On semiparametric transformation cure models. *Biometrika*. 2004; 91:331.
21. Luo X, Boyett J. Estimation of a threshold parameter in Cox regression. *Communications in Statistics: Theory and Methods*. 1997; 26:2329–2346.
22. Martinussen T, Scheike T. Aalen additive hazards change-point model. *Biometrika*. 2007; 94:861.
23. Matthews D, Farewell V. On testing for a constant hazard against a change-point alternative. *Biometrics*. 1982; 38:463–468. [PubMed: 7052152]
24. Merrill R, Bird J. Effect of young age on prostate cancer survival: a population-based assessment (United States). *Cancer Causes and Control*. 2002; 13:435–443. [PubMed: 12146848]
25. Nguyen H, Rogers G, Walker E. Estimation in change-point hazard rate models. *Biometrika*. 1984; 71:299–304.
26. Optenberg S, Thompson I, Friedrichs P, Wojcik B, Stein C, Kramer B. Race, treatment, and long-term survival from prostate cancer in an equal-access medical care delivery system. *Jama*. 1995; 274:1599. [PubMed: 7474244]
27. Peng Y, Dear K. A nonparametric mixture model for cure rate estimation. *Biometrics*. 2000; 56:237–243. [PubMed: 10783801]
28. Peters N, Armstrong K. Racial differences in prostate cancer treatment outcomes: a systematic review. *Cancer nursing*. 2005; 28:108. [PubMed: 15815180]
29. Pons O. Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *Ann Statist*. 2003; 31:442–463.
30. Sy J, Taylor J. Estimation in a Cox proportional hazards cure model. *Biometrics*. 2000; 56:227–236. [PubMed: 10783800]
31. Tsodikov A, Ibrahim J, Yakovlev A. Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*. 2003; 98:1063–1079. [PubMed: 21151838]
32. Underwood W, DeMONNER S, Ubel P, Fagerlin A, Sanda M, Wei J. Racial/ethnic disparities in the treatment of localized/regional prostate cancer. *The Journal of urology*. 2004; 171:1504–1507. [PubMed: 15017208]
33. Ward E, Jemal A, Cokkinides V, Singh G, Cardinez C, Ghafoor A, Thun M. Cancer disparities by race/ethnicity and socioeconomic status. *CA: a cancer journal for clinicians*. 2004; 54:78. [PubMed: 15061598]
34. Worsley K. Exact percentage points of the likelihood-ratio test for a change-point hazard-rate model. *Biometrics*. 1988; 44:259–263. [PubMed: 3358993]

Appendix A. Variance Formulae

Define $f_j(\theta, t) = (1 - c_j)\lambda_j \exp(-\lambda_j t)$ and $S_j(\theta, t) = (1 - c_j) \exp(-\lambda_j t) + c_j$ for $j = 1, 2$ and let $Z_i = (\tilde{T}_i, \Delta_i)$ for $i = 1 \dots n$. Write the censored data log-likelihood corresponding to $f_j(\theta, t)$ and $S_j(\theta, t)$ as

$$l_j(\theta, z) = \Delta \log f_j(\theta, t) + (1 - \Delta) \log S_j(\theta, t) \text{ for } j=1, 2.$$

The averaged log-likelihood for (3) is written

$$l_n(\theta) = n^{-1} \sum_{i=1}^n [K_n(\tau - X_i)l_1(\theta, Z_i) + \{1 - K_n(\tau - X_i)\}l_2(\theta, Z_i)],$$

while the corresponding non-smooth averaged log-likelihood is

$$l_n^*(\theta) = n^{-1} \sum_{i=1}^n [I(X_i \leq \tau)l_1(\theta, Z_i) + I(X_i > \tau)l_2(\theta, Z_i)].$$

Write the score equations for $I_n(\theta)$ with respect to θ as

$$R_n(\theta) = (R_{nc_1}(\theta), R_{n\lambda_1}(\theta), R_{nc_2}(\theta), R_{n\lambda_2}(\theta), R_{n\tau}(\theta)) = (R_{n\varphi}(\theta), R_{n\tau}(\theta)).$$

where for $j = 1, 2$:

$$\begin{aligned} R_{nc_j}(\theta) &= n^{-1} \sum_{i=1}^n \{[(2-j)K_n(\tau - X_i) + (j-1)\{1 - K_n(\tau - X_i)\}][-\Delta_i/(1 - c_j) + (1 - \Delta_i)(1 - \exp(-\lambda_j \tilde{T}_i))/S_j(\theta, \tilde{T}_i)]\} \\ R_{n\lambda_j}(\theta) &= n^{-1} \sum_{i=1}^n \{[(2-j)K_n(\tau - X_i) + (j-1)\{1 - K_n(\tau - X_i)\}][\Delta_i(\lambda_j^{-1} - \tilde{T}_i) - (1 - \Delta_i)(1 - c_j)\tilde{T}_i \exp(-\lambda_j \tilde{T}_i)/S_j(\theta, \tilde{T}_i)]\} \\ R_{n\tau}(\theta) &= n^{-1} \sum_{i=1}^n \dot{K}_n(\tau - X_i)[l_1(\theta, \tilde{T}_i) - l_2(\theta, \tilde{T}_i)]. \end{aligned}$$

Let $Q_n(\theta) = R_n(\theta)/\theta$. Partition $Q(\theta)^{-1}$ as:

$$\begin{bmatrix} Q(\theta)_{\varphi}^{-1} & Q(\theta)_0^{-1} \\ Q(\theta)_0^{-1} & Q(\theta)_{\tau}^{-1} \end{bmatrix}.$$

Formulae for the relevant elements of $Q_n(\theta)$ are:

$$\begin{aligned} Q_{nc_j c_j}(\theta) &= \partial R_{nc_j}(\theta) / \partial c_j = \\ & n^{-1} \sum_{i=1}^n \{[(2-j)K_n(\tau - X_i) + (j-1)\{1 - K_n(\tau - X_i)\}] \\ & [-\Delta_i(1 - c_j)^{-2} - (1 - \Delta_i)(1 - \exp(-\lambda_j \tilde{T}_i))^2 / S_j(\theta, \tilde{T}_i)^2]\} \\ Q_{n\lambda_j \lambda_j}(\theta) &= \partial R_{n\lambda_j}(\theta) / \partial \lambda_j = \\ & n^{-1} \sum_{i=1}^n \{[(2-j)K_n(\tau - X_i) + (j-1)\{1 - K_n(\tau - X_i)\}] \\ & [-\Delta_i \lambda_j^{-2} - (1 - \Delta_i)(1 - c_j)\tilde{T}_i^2 [-\exp(-\lambda_j \tilde{T}_i) / S_j(\theta, \tilde{T}_i) + \\ & (1 - c_j) \exp(-2\lambda_j \tilde{T}_i) / S_j(\theta, \tilde{T}_i)^2]]\} \\ Q_{n\lambda_j c_j}(\theta) &= \partial R_{n\lambda_j}(\theta) / \partial c_j = \\ & n^{-1} \sum_{i=1}^n K_n(\tau - X_i)(1 - \Delta_i)[\tilde{T}_i \exp(-\lambda_j \tilde{T}_i) / S_j(\theta, \tilde{T}_i) \\ & + (1 - \exp(-\lambda_j \tilde{T}_i))\tilde{T}_i(1 - c_j) \exp(\lambda_j \tilde{T}_i) / S_j(\theta, \tilde{T}_i)^2] \\ Q_{n\tau \tau}(\theta) &= \partial R_{n\tau}(\theta) / \partial \tau = n^{-1} \sum_{i=1}^n \ddot{K}_n(\tau - X_i)[l_1(\theta, \tilde{T}_i) - l_2(\theta, \tilde{T}_i)]. \end{aligned}$$

Define $\Sigma_\varphi = \text{Cov}(R_{n\varphi})$ and $\Sigma_\tau = \lim_n V\{h_n R_n(\tau)\}$. Then $V(\widehat{\varphi}) = Q(\theta_0)_\varphi^{-1} \sum_\varphi Q(\theta_0)_\varphi^{-1}$ and $V(\widehat{\tau}) = Q(\theta_0)_\tau^{-2} \sum_\tau$.

Figure 1.
Kaplan-Meier curve for prostate cancer survival for the SEER cohort

Figure 2.

Kaplan-Meier and model-based survival plots. Kaplan-Meier plots are shown with solid lines and plots based on our exponential cure model are shown with dashed lines

Table A1

Simulation results

		$X \sim \text{tN}(1, 1, 0, 2)$				$X \sim \text{Uniform}(0, 1)$			
	θ_0	Est	SE_{mc}	SE	θ_0	Est	SE_{mc}	SE	
N = 250									
c_1	0.4	0.395	0.075	0.065	0.4	0.383	0.059	0.057	
λ_1	0.3	0.318	0.067	0.058	0.3	0.328	0.055	0.053	
c_2	0.3	0.313	0.056	0.055	0.3	0.320	0.052	0.051	
λ_2	0.4	0.398	0.068	0.060	0.4	0.389	0.059	0.054	
τ	1	0.989	0.287	0.182	0.5	0.499	0.136	0.166	
N = 500									
c_1	0.4	0.392	0.046	0.046	0.4	0.381	0.046	0.042	
λ_1	0.3	0.312	0.045	0.041	0.3	0.319	0.039	0.038	
c_2	0.3	0.307	0.041	0.039	0.3	0.315	0.038	0.037	
λ_2	0.4	0.396	0.046	0.043	0.4	0.386	0.040	0.039	
τ	1	0.997	0.177	0.131	0.5	0.498	0.100	0.104	
N = 1000									
c_1	0.4	0.392	0.034	0.032	0.4	0.399	0.031	0.031	
λ_1	0.3	0.309	0.031	0.029	0.3	0.316	0.030	0.028	
c_2	0.3	0.305	0.028	0.028	0.3	0.314	0.026	0.027	
λ_2	0.4	0.395	0.031	0.030	0.4	0.391	0.030	0.029	
τ	1	1.003	0.123	0.090	0.5	0.501	0.060	0.063	
N = 3000									
c_1	0.4	0.395	0.020	0.019	0.4	0.390	0.019	0.018	
λ_1	0.3	0.306	0.018	0.017	0.3	0.309	0.017	0.016	
c_2	0.3	0.305	0.017	0.016	0.3	0.309	0.016	0.016	
λ_2	0.4	0.396	0.018	0.018	0.4	0.392	0.018	0.017	
τ	1.0	1.001	0.044	0.042	0.5	0.499	0.031	0.030	

θ_0 = true values of parameters

Est = mean of estimates

SE_{mc} = Monte Carlo standard error

SE = analytic standard errors

Table A2

Parameter Estimates and Confidence Intervals for SEER data

Parameter	Black		White	
	Estimate	95% CI	Estimate	95% CI
Change-point estimate ($\hat{\tau}$)	71	(68, 75)	72	(71, 73)
Cure estimate for age $\leq \hat{\tau}$	0.37	(0.28, 0.44)	0.47	(0.45, 0.49)
Hazard estimate for age $\leq \hat{\tau}$	0.49	(0.40, 0.57)	0.44	(0.42, 0.47)
Cure estimate for age $> \hat{\tau}$	0.32	(0.26, 0.37)	0.39	(0.37, 0.41)
Hazard estimate for age $> \hat{\tau}$	0.55	(0.48, 0.62)	0.54	(0.52, 0.56)