Published in final edited form as:
*Comput Stat Data Anal*. 2009 October 1; 53(12): 4073–40. doi:10.1016/j.csda.2009.04.011.

# Semiparametric Bayesian approaches to joinpoint regression for population-based cancer survival data

**Pulak Ghosh**[a,*,1], **Lan Huang**[b,2], **Binbing Yu**[c,3], and **Ram C. Tiwari**[d,4]

[a]Department of Mathematics and Statistics, Georgia State University, Atlanta, USA

[b]Statistical Research and Applications Branch, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD 20892-8317, USA

[c]National Institute of Aging, Bethesda, MD 20892-8317, USA

[d]Office of Biostatistics, Food and Drug Administration, Silver Spring, MD, USA

## Abstract

According to the American Cancer Society report (1999), cancer surpasses heart disease as the leading cause of death in the United States of America (USA) for people of age less than 85. Thus, medical research in cancer is an important public health interest. Understanding how medical improvements are affecting cancer incidence, mortality and survival is critical for effective cancer control. In this paper, we study the cancer survival trend on the population level cancer data. In particular, we develop a parametric Bayesian joinpoint regression model based on a Poisson distribution for the relative survival. To avoid identifying the cause of death, we only conduct analysis based on the relative survival. The method is further extended to the semiparametric Bayesian joinpoint regression models wherein the parametric distributional assumptions of the joinpoint regression models are relaxed by modeling the distribution of regression slopes using Dirichlet process mixtures. We also consider the effect of adding covariates of interest in the joinpoint model. Three model selection criteria, namely, the conditional predictive ordinate (CPO), the expected predictive deviance (EPD), and the deviance information criteria (DIC), are used to select the number of joinpoints. We analyze the grouped survival data for distant testicular cancer from the Surveillance, Epidemiology, and End Results (SEER) Program using these Bayesian models.

## 1. Introduction

The fight against cancer, escalated in the early 1970s with the introduction of the National Cancer Act during the presidency of Richard M. Nixon, has brought dramatic improvements in prevention, screening, and treatment that have had a major impact on our ability to reduce the cancer burden in the USA. The study of cancer trends in terms of all related measures associated with cancer (i.e., cancer mortality, incidence, prevalence and survival) is very helpful in understanding the impact and effectiveness of all our efforts on extending the life of cancer patients and reducing the occurrence of new cancer and cancer deaths.



*Corresponding author. Tel.: +1 404 413 6435, pulakghosh@gmail.com (P. Ghosh).
[1]Associate Professor.
[2]Mathematical Statistician.
[3]Statistician.
[4]Associate Director.

Joinpoint models (Kim et al., 2000) have recently been used to model the progress and trend of cancer mortality rates, which not only provide the varying trend information, but also indicate the significant time points at which the measures experience a promising improvement that may relate to earlier detection or better treatment. Such models can also be used for incidence data to evaluate the trend of the cancer burden (Ries et al., 2006). However, analyzing cancer incidence and mortality is not always enough to understand the benefits of medical breakthroughs in cancer as it does not provide information on the situation of the patients during their lifetime after diagnosis.

To better understand life after diagnosis, we use the net survival rate (Cronin and Feuer, 2000), which is a key measure for the population to assess the chance of cancer survival after diagnosis till the occurrence of death due to cancer. The survival time is usually defined as the time from diagnosis to death. Assuming that a person may survive for many more years after being diagnosed with cancer, information on survival rates can play an important role in planning treatment strategies. In addition, differences in survival rates between defined subgroups of patients allow clinicians and policy makers to better target interventions. The survival trend may not have such a big increasing or decreasing pattern as we observe for incidence or mortality, but as discussed in Feuer et al. (1991), the survival rate usually improves dramatically after the introduction of an effective treatment, and then levels off after the dissemination of the cancer treatment has been fully realized to the population, thus indicating a possibility of the presence of multiple numbers of change points in survival function. Here, we consider incorporating a joinpoint model into the survival model for capturing possible big changes in survival trend. Besides treatment, survival may also be affected by the introduction and dissemination of new screening techniques and prevention activities. Therefore, it is essential to model the trend of survival at the population level to understand the change in survival patterns over time and to assess the effort of the whole country in improving the survival chance and extending the life of cancer patients.

There are three national cancer data sources in the USA, namely, the national cancer data base under the Commission on Cancer of American College of Surgeons, the Surveillance, Epidemiology, and End Results (SEER) program under the National Cancer Institute (NCI), and the National Program of Cancer Registries (NPCR) under the Centers for Disease Controls (CDC). However, the national cancer database is actually hospital based, not really population based, and the NPCR does not require the collection of survival data. The SEER data (1973–current) is the unique source of population-based data for providing a good cancer survival trend estimation and for identifying the points of dramatic changes in the survival trend at the population level with long-term follow-up. Starting from 1973–75, the SEER program included 9 registries covering almost 10% of the USA population (SEER 9 registries), and then went through several expansions in the early 1990s and 2000 (SEER 13 and SEER 17 registries). Currently, there are 17 registries and about 26% of the total USA population is included in the areas covered by the SEER 17 registries, excluding Alaska.

The SEER population data is not very precise in terms of patients' personal information because of confidentiality issues. Specifically, the SEER program does not disclose the exact survival time for each subject, but the survival months. The SEER program provides the number of patients alive at the beginning of a time period (month or year), the observed number of deaths during this period, and the cause of their death. Therefore, we do not have exact individual survival data, but the grouped survival data. To capture the change of trend in survival, instead of overall survival improvement, we model the population survival trend with joinpoints to analyze the SEER grouped survival data.

In this article, we consider a Bayesian approach to joinpoint regression for population-based cancer survival data. In particular, we use a Poisson regression model (Frome, 1983; Raftery and Akman, 1986) for the number of deaths, by intervals, as it is more appropriate for rare cancer sites with a small number of deaths or for cancer sites with good survival. Our model also incorporates the covariates. The inclusion of covariates can change both the number and the location of joinpoints, as we show in the analysis of our data. Thus, we build a more flexible model that can easily accommodate multiple covariates. While Joinpoint software (http://srab.cancer.gov/joinpoint/) assumes that the joinpoints occur on the discrete time grid and searches for the joinpoints using a grid search method (Lerman, 1980), we allow an option for implementing a discrete or a continuous prior (see, for example, Carlin et al. (1992)) distribution on the locations of the multiple joinpoints under a Bayesian paradigm. These priors also allow the user to impose prespecified minimum gaps in between two consecutive joinpoints. The confidence intervals for the joinpoints derived using the Joinpoint software are based on asymptotic results and are reported in Lerman (1980) and Feder (1975). These asymptotic distributional results may not hold if the "true" model contains a smaller number of joinpoints than the fitted model. The proposed Bayesian method not only overcomes this drawback, but also provides a measure of uncertainty related to the number and locations of joinpoints in the data. In addition, the parametric distributional assumptions for the slope parameters of the relative survival model are relaxed by using a semiparametric Bayesian method, wherein, instead of assuming the slopes to be normally distributed, they are assumed to follow a mixture of normals. The class of mixing distributions proposed is quite large. In particular, we assume a Dirichlet process (DP) prior (Ferguson, 1973; Antoniak, 1974) on the mean of the normal distribution for the slopes, resulting in a DP mixture (DPM) prior. For the model selection, to choose from the $(K + 1)$ models, $M_0, M_1, \ldots, M_K$, corresponding to no joinpoints, one joinpoints, …, up to $K$ joinpoints, we use three model selection criteria, namely, the conditional predictive ordinate (CPO; Chen et al., 2000, Chapter 10), the deviance information criterion (DIC; Spiegelhalter et al., 2002), and the expected predictive deviance (EPD) based on a posterior predictive distribution (Laud and Ibrahim, 1996).

The rest of the paper is organized as follows. In Section 2, the joinpoint survival model for grouped survival data is introduced. The priors for the parameters in the joinpoint survival model are discussed in Section 3. The measures for model selection are included in Section 4. In Section 5, we illustrate the application of the parametric and semiparametric Bayesian joinpoint models on the testicular cancer survival data from the SEER database using the Markov Chain Monte Carlo (MCMC) methodology. This article ends with a discussion in Section 6.

## 2. Joinpoint survival models for group survival data

Let $T$ denote the survival time, $x$ the year of diagnosis, and $z$ the vector of other covariates, such as race, age, and sex, etc. The change in survival trend could occur at any calendar time $x$. We assume a proportional hazards model for modeling the hazard for surviving after diagnosis, and thus the hazard function for surviving $t$ years since diagnosis for persons diagnosed with cancer is modeled as follows:

$$\lambda(t|x, z) = \lambda_0(t) \exp\{h(x, z)\}. \quad (1)$$

Here $\lambda_0(t)$ is the baseline hazard, and

$$h(x,z)=\beta x+\sum_{k=1}^{K}\delta_k(x-\tau_k)^+ +\gamma' z \tag{2}$$

indicates the trend in hazard with respect to calendar time $x$ and other covariates of interest $z$, and $(x-\tau_k)^+ = x-\tau_k$ if $x-\tau_k > 0$, and 0 otherwise. We assume that the maximum number of joinpoints in the model is a known finite number, $K$. We write the parameter vector as $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \beta, \boldsymbol{\delta}, \boldsymbol{\tau})$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_J)$, $\beta$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_K)$ and $\tau = (\tau_1, \ldots, \tau_K)$ are the covariate parameters, the slope parameters, and the joinpoints, respectively. The survival model, with $h(x,z)$ defined as in (2), is referred to as a $K$-joinpoint survival model (Zelterman et al., 1994; Luo et al., 1997), as there are $K+1$ segments, with the slope coefficient $\beta_1 = \beta$, and $\beta_k=\beta+\sum_{l=1}^{k-1}\delta_l$, $k = 2, 3, \ldots, K+1$ (see, for example, Zelterman et al. (1994), Luo et al. (1997) and Chang and Huang (1997)). It is important to note that the joinpoints are introduced, not in survival time $t$, but in the year of diagnosis $x$, and they represent the relative change in the hazard at $t$ with respect to $x$.

In the SEER program, individual survival time after diagnosis is not available. Instead, as mentioned earlier, the survival times after diagnosis are grouped into intervals $I_j = [t_{j-1}, t_j)$, $j = 1, 2, \ldots, J$, where $t_0 = 0$, and $t_j = J$ is the end of follow-up. The lengths of the intervals are defined as one-year, the event as the death due to cancer of interest, and the people dying from other causes or lost to follow-up are considered as censored. For the patient cohort diagnosed in year $x$ with age $z$, let $n_{xzj}$ be the number of people alive at the beginning of interval $I_j$, $d_{xzj}$ be the number of cancer deaths, and $l_{xzj}$ be the number of patients lost to follow-up or dying from other causes in interval $I_j$. Following Gail (1975), the adjusted number of person-years at risk is $r_{xzj}=n_{xzj}-\frac{1}{2}l_{xzj}$.

A binomial distribution can be used to model the number of deaths from the cancer of interest in each interval (i.e., year). However, the binomial distribution may not be an appropriate assumption for rare cancers with a smaller number of deaths or cancer sites with good survival. Sometimes, there are even no deaths in certain intervals. In this scenario, we advocate the use of a Poisson distribution instead of a binomial distribution for the observed number of deaths, i.e., we assume that the number of cancer deaths in $I_j$ follows a Poisson distribution, $d_{xzj} \sim Poi(r_{xzj} \times \lambda_j(x,z))$. Note that $\lambda j(x,z)$ is the death rate during the interval $I_j$ given that a patient is alive at the beginning of the interval $I_j$, and is given by

$$\lambda_j(x,z)=P(T<t_j|T\geq t_{j-1};x,z)=1-\frac{S(t_j|x,z)}{S(t_{j-1}|x,z)},\ j=1,\ldots,J, \tag{3}$$

and $S(t|x,z) = S_0(t)^{\exp\{h(x,z)\}}$ is the survival function under the proportional hazards assumption, with $S_0(t)$ as the baseline survival function.

The likelihood function for the grouped survival data $Y = \{x, z, (r_{xzj}, d_{xzj}), j = 1, \ldots, J\}$, is then

$$L(\theta|Y)=\prod_x\prod_z\prod_{j=1}^{J}\frac{1}{d_{xzj}!}(r_{xzj}\times\lambda_j(x,z))^{d_{xzj}}\exp(-r_{xzj}\times\lambda_j(x,z)). \tag{4}$$

To estimate cause-specific survival, we require accurate information on the cause of death. However, the information on cause of death is not always available in the population data base. Thus, when the event of interest is death due to a certain disease and the cause of death is not known, it is not possible to accurately estimate the number of cancer deaths due to the disease in question. Therefore, we may not have accurate values of $d_{xzj}$ in Eq. (4). In such situations, the relative survival rate or ratio (Edere et al., 1961), defined as the observed survival rate in patients for a specified time interval divided by the expected survival rate in the same time interval in a healthy population free of the cancer of interest, is an alternative way to estimate the cancer net survival. In relative survival analysis, $d_{xzj}$ is defined as the number of patients dying from all causes instead of the cancer of interest. $l_{xzj}$ becomes the number of patients lost to follow-up during interval $I_j$. We then assume that

$$d_{xzj} \sim Poi(r_{xzj} \times [1 - (1 - \lambda_j(x, z)) \times E_j(x, z)]),$$

where $(1 - \lambda_j(x, z))$ is the interval relative survival probability and $E_j(x, z)$ is the expected probability of surviving interval $I_j$ for a healthy population that can be obtained from the life tables for the general population (National Center for Health Statistics, 2003). Thus $(1 - \lambda_j(x, z))E_j(x, z)$ is the overall survival probability for interval $I_j$. The likelihood for the relative survival analysis, given $D = \{x, z, r_{xzj}, d_{xzj}, E_j(x, z), j = 1, \ldots, J\}$, is then

$$L(\theta|D)=\prod_x\prod_z\prod_{j=1}^{J}\frac{1}{d_{xzj}!}(r_{xzj} \times [1 - (1 - \lambda_j(x, z)) \times E_j(x, z)])^{d_{xzj}}\exp(r_{xzj} \times [1-(1-\lambda_j(x, z)) \times E_j(x, z)]). \quad (5)$$

From Eqs. (1) and (5), we have

$$\log\{-\log[1 - \lambda_j(x, z)]\}=\log\left\{-\log\left[\frac{S_0(t_j)}{S_0(t_{j-1})}\right]\right\}+h(x, z). \quad (6)$$

Let $\alpha_j=\log\{-\log[\frac{S_0(t_j)}{S_0(t_{j-1})}]\}$. Then, the baseline survival function can be expressed as $S_0(t)=\exp[-\sum_{l=0}^{j}e^{\alpha_l}]$.

In the above set up, $z$ is a vector of multiple covariates; however, in our following analysis, we treat it as single covariate age associated with the parameter $\gamma$. The annual percentage rate (APC) in death rates in the $k$th segment is $APC_k = [\exp(\beta_k) - 1]100\%$. A negative value of APC implies that the death rates $\lambda_j(x, z)$ decrease as $x$ increases.

## 3. Prior distribution

Let $\pi(\boldsymbol{\theta})$ be the joint prior distribution of $\boldsymbol{\theta}$ and let $L(\boldsymbol{\theta}|Y)$ be the likelihood function given by (4) for survival analysis. The joint posterior distribution of $\boldsymbol{\theta}$ is

$$\pi(\theta|Y) \propto L(\theta|Y) \times \pi(\theta).$$

We assume that the prior distributions for the parameters are mutually independent, i.e.,

$$\pi(\theta)=\pi(\beta)\left\{\prod_{k=1}^{K}\pi(\delta_k)\right\}\pi(\tau_1,\ldots,\tau_K)\pi(\gamma).$$

In particular, we assume conjugate priors for the model parameters:

$\pi(\gamma_j)\sim N(\gamma_{0j},\sigma^2_{\gamma_j})$, $j=1,2,\ldots,J$;$\pi(\beta)\sim N(\beta_0,\beta_0)$.

In addition to the specification of prior probability distribution for **θ**, we also assume that the hyperparameters ($\gamma_{0j}$, $\delta_{0k}$, $A_{0j}$, $d_{0k}$) are independent and have the following priors:

$\gamma_{0j} \sim N(\mu_\alpha, \tau_\alpha)$; $\delta_{0k} \sim N(\mu_\delta, \tau_\delta)$; $A_{0j} \sim IG(c_1, d_1)$; $d_{0k} \sim IG(c_2, d_2)$, where $IG(c, d)$ denotes an inverse gamma distribution with shape parameter $c$ and scale parameter $d$.

The prior specification of the random slopes $\delta_k$, $k = 1, 2, \ldots, K$, is an important feature of the model since it may crucially impair or influence the accurate estimation of the random slopes associated with the joinpoint. Although the usual prior is the normal distribution for the random slope $\delta_k$, a normal distribution has limited flexibility because it is unimodal, thin tailed, and fails to accommodate skewness. Thus, to bring more flexibility in modeling the distribution for the random effects (i.e., the slopes $\delta_k$ at the joinpoints), we propose a robust distribution to construct the priors for $\delta_k$ by using a Dirichlet Process (DP) prior (Escober, 1994; Escober and West, 1995; MacEachern, 1994; Ishwaran and James, 2001)). Thus we assume the following:

$$\delta_k \overset{\text{iid}}{\sim} \pi_k I[\delta_k=0]+(1-\pi_k)G \tag{7}$$

$$G \overset{\text{ind}}{\sim} DP(\nu, G_0) \tag{8}$$

where $I$ is the indicator function, $\nu \geq 0$ is a scalar precision parameter and $G_0$ is a parametric baseline distribution. The above model can be better explained by the fact that the measure $G_0$ can be thought of as a prior guess of the distribution of random effects and ν as a measure of the strength of this belief (Ferguson, 1973; Sethuraman and Tiwari, 1982; Basu and Tiwari, 1982). Large values of ν lead to a $G$ that is very close to $G_0$. Small values of ν allow $G$ to deviate more from $G_0$ and put most of its probability mass on just a few atoms.

The above formulation of the slope has three important implications: (a) it assigns positive prior probabilities at 0 for each $\delta_k$, to permit assessment of hypotheses that fewer slopes are actually nonzero, and hence to infer values of the number of significant nonzero slopes, (b) it better accommodates the lack of knowledge of the distributional structure by allowing a richer class of distributions (compared to a parametric family) and (c) it gives positive probability to allow the nonzero $\delta_k$ to cluster.

There are several ways to implement a DP prior. Recent research has focused on using the following constructive definition of the DP (Sethuraman and Tiwari, 1982; Sethuraman, 1994) to produce MCMC algorithms:

$$G(\cdot)=\sum_{r=1}^{\infty} p_r \delta_{Z_r}(\cdot), \quad \text{where } Z_r \overset{\text{iid}}{\sim} G_0(\cdot), \quad r=1,\ldots,\text{and}$$
$$\text{with } p_1=V_1, p_r=V_l\prod_{j=1}^{r-1}(1-V_j), \quad r=2,\ldots, \text{ and } V_l \overset{\text{iid}}{\sim} \text{Beta}(1,\ \nu), l \geq 1 \tag{9}$$

for $r = 1, 2, \ldots, \infty$, with $\delta_Z$ denoting the degenerate distribution with all its mass at $Z$. Because the infinite series in (9) is almost surely convergent, the random vector $(p_r, Z_r)$, as $r$ increases to infinity, will have a diminishing effect on the prior distribution and thus on the posterior distribution of $\delta_k$. Thus, in practice one can truncate the above mixture at some large $R \sum_{r=1}^{R} p_r = 1$, and thus after the truncation $G$ can be represented as

$$G \approx \sum_{r=1}^{R} p_r \delta_{Z_r}. \tag{10}$$

The advantage of this approximation is that the model reduces to a finite mixture model (Ghosh and Rosner, 2007; Ohlssen et al., 2007) and can be fitted using the standard MCMC methods and implemented in the freely available WinBUGS software (Spiegelhalter et al., 2005).

The full Bayesian model in the present context is completed by assigning prior distributions for the DPM parameters ν and $G_0$. It is assumed that

$$\nu \sim \text{Gamma}(\nu_1, \nu_2), G_0|\eta \sim N(0, \eta^2), \eta^{-2} \sim IG(\eta_1, \eta_2), \pi_k \sim \text{Beta}(a, b). \tag{11}$$

The hyperparameters of all the above prior distributions are assumed to be known.

### 3.1. Prior for joinpoints

Thus, we assume two prior distributions for the multiple joinpoints: one is a discrete prior (Tiwari et al., 2005) and the other is a continuous prior (Ghosh et al., 2009).

**3.1.1. Discrete prior—**Let $x_1, \ldots, x_m$ denote the diagnosis years in the observed data, and let $K \ll m$ be the prespecified number of joinpoints. In the discrete case (Tiwari et al., 2005), the prior for the joinpoints $\pi(\tau_1, \ldots, \tau_K)$ is the product of

$$\pi(\tau_1) \propto \frac{1}{m-(2l+K-1)}, \tau_1 \in \{x_{l+1}, \ldots, x_{m-l-K+1}\}, l \geq 0$$
$$\pi(\tau_u|\tau_{u-1}=x_{l'}) \propto \frac{1}{m-(2l+l'+K-u)}, \tau_u \in \{x_{l'+l+1}, \ldots, x_{m-l-K+u}\},\ u=2,\ldots,K,$$

where $x_1, x_2, \ldots, x_m$ denote the observed values of the covariate $x$ (the year of diagnosis). Note that the distribution of joinpoint $\tau_1$ is a discrete uniform on $\{x_{l+1}, \ldots, x_{m-l-k+1}\}$, leaving out $l(\geq 0)$ values of $x$ at both ends, and the conditional distribution of $\tau_u$; given $\{\tau_{u-1} = x_{l'}\}$ is also a discrete uniform distribution on $\{x_{l'+l+1}, \ldots, x_{m-l-K+u}\}$, $u = 2, \ldots, K$, leaving out $l$ data points at both ends.

**3.1.2. Continuous prior—**Sometimes a continuous prior may be more appropriate, as it allows the joinpoints to occur anywhere on the continuous scale between the first and last

observed diagnosis years $x_1$ and $x_m$, inclusively. Thus, assuming a continuous prior for joinpoints may give a more accurate estimates of the joinpoints. Following Ghosh et al. (2009), we define a continuous prior for joinpoints, by modeling the spacings or gaps among the joinpoints instead of the joinpoints themselves. We consider the spacings $\tau_1 - x_1$, $\tau_2 - \tau_1$, …, $\tau_K - \tau_{K-1}$, $x_m - \tau_K$ and normalize these spacings by dividing by the range $x_m - x_1$. The normalized spacings then provide a partition of [0, 1] with a Dirichlet distribution (Willks, 1962) as a natural prior. This is given by

$$\frac{1}{x_m - x_1}(\tau_1 - t_1, \tau_2 - \tau_1, \ldots, \tau_K - \tau_{K-1}, x_m - \tau_K) \sim \text{Dirichlet}(a_1, a_2, \ldots, a_{k+1}). \tag{12}$$

This Dirichlet distribution can also be expressed as a vector of $K + 1$ independent Gamma random variables with a common shape parameter and (possibly) different scale parameters $a_j$, $j = 1, 2, \ldots, K + 1$, normalized by their sum; this characterization can be useful in Markov chain sampling. Finally, when $a_j = 1$, $j = 1, \ldots, K + 1$, the resulting Dirichlet prior is an uniform distribution on the simplex: a simple analogue to the uniform prior construction considered by Tiwari et al. (2005) in the discrete case.

We can easily modify the above spacings prior to allow for some prespecified gaps. Thus, one may decide to allow a gap of at least $\varpi_j$ between $\tau_{j-1}$ and $\tau_j$, $j = 1, \ldots, K + 1$ (here $\tau_0 = t_1$ and $\tau_{K+1} = t_n$). This can be easily incorporated within our proposed spacings prior simply by replacing (12) with

$$\frac{1}{x_m + x_1 - \sum_{u=1}^{K+1} \varpi_u}(\tau_1 - t_1 - \varpi_1,\ \tau_2 - \tau_1 - \varpi_2, \ldots, \tau_K - \tau_{K-1} - \varpi_L,\ x_m - \tau_K - \varpi_{K+1}) \sim \text{Dirichlet}(a_1, a_2, \ldots, a_{K+1}). \tag{13}$$

Similar priors have been considered in Bayesian piecewise regression models in other contexts; see Green (1995), Denison et al. (1998), and Kass et al. (2003).

## 4. Model selection and model adequacy

In the joinpoint regression model, each joinpoint indicates a change in the underlying slope, and often these changes are one of the primary interest of the analysis. Thus, a crucial issue in joinpoint model is the selection of the number of joinpoints. More specifically, consider the collection of $(K + 1)$ models $[M_0, M_1, \ldots, M_K]$, where $M_k$, $0 \le k \le K$, for known $K$, refers to the joinpoint regression model in (2) with exactly $k$ joinpoints ($M_0$ corresponds to the simple regression model with no joinpoints). The selection of the number of joinpoints then amounts to selecting a model from these $(K + 1)$ choices.

We present three popular approaches to the model selection in our Bayesian analysis, namely, the conditional predictive ordinate (CPO) (Gelfand et al., 1992; Chen et al., 2000), the expected predictive deviance (EPD) (Laud and Ibrahim, 1996; Gelfand and Ghosh, 1998) and the Deviance information criterion (DIC) (Spiegelhalter et al., 2002).

The CPO statistic is a very useful model assessment tool which has been widely used in the literature in various contexts. For a detailed discussion of the CPO statistic and its applications to model assessment, see Chen et al. (2000, Chapter, 10).

For the *i*th observation, the CPO statistic under model $M_k$ is defined as

$$\mathrm{CPO}_i = p(y_i|\mathbf{y}^n_{(-i)}) = E\{f(y_i|\theta_k)|\mathbf{y}^n_{(-i)}\}, \tag{14}$$

where $\mathbf{y}^n_{(-i)}$ denotes the rest of the data after deleting the *i*th observation, $\boldsymbol{\theta}_k$ is the set of parameters of the model $M_k$ and $f(y_i|\boldsymbol{\theta}_k)$ is the sampling density of the model evaluated at the *i*th observation. The expectation above is taken with respect to the posterior distribution of the model parameter $\boldsymbol{\theta}_k$ given the cross-validated data $\mathbf{y}^n_{(-i)}$. The $\mathrm{CPO}_i$ can be computed from the MCMC samples drawn from the posterior by using the simplification

$$\mathrm{CPO}_i = \left( \frac{1}{N} \sum_{m=1}^{N} \frac{1}{f(y_i|\theta_k^{(m)})} \right)^{-1},$$

where *N* is the number of simulations. Thus, CPO*i* can be interpreted as the height of this marginal density or probability at $y_i$. Large values of $\mathrm{CPO}_i$ imply a better fit of the model. A useful summary statistic of the $\mathrm{CPO}_i$s is the logarithm of the pseudomarginal likelihood, LPML, defined as

$$\mathrm{LPML} = \sum_{i=1}^{n} \log(CPO_i).$$

The model with larger LPML value is the better fitting model. Note that the LPML is always well defined as long as the posterior predictive density is proper. Thus, the LPML is well defined even under improper priors. Additionally it is computationally very stable.

Another summary measure for model selection is to use the predictive performance criterion proposed by Laud and Ibrahim (1996) and Gelfand and Ghosh (1998). Given a finite number of models, the criterion is based on the predictive performances of the models. Let ypred be a replicate of the observed data vector $y_{\mathrm{obs}}$. The posterior predictive distribution of $y_{pred}$ under model $M_k$ is

$$f^{(k)}(y_{\mathrm{pred}}|y_{\mathrm{obs}}) = \int f(y_{\mathrm{pred}}|\theta_k) f(\theta_k|y_{\mathrm{obs}})\, \mathrm{d}\theta_k \tag{15}$$

where $\theta_k$ denotes the set of parameters under model $M_k$, $f(\theta_k|y_{\mathrm{obs}})$ is the posterior density and $f(y_{\mathrm{pred}}|\theta_k)$ is the density of the predicted value. The model selection criterion called the expected predictive deviance (EPD) chooses the model *M* with smallest value of

$$E^{(M)}[d(y_{\mathrm{pred}}, y_{\mathrm{obs}})|y_{\mathrm{obs}}]$$

where $d(y_{\mathrm{pred}}, y_{\mathrm{obs}})$ is a discrepancy function and the expectation is with respect to the predictive distribution (15). We take $d(y_{\mathrm{pred}}, y_{\mathrm{obs}}) = \|y_{\mathrm{pred}} - y_{\mathrm{obs}}\|^2$, where $\|x\|^2$ denotes the sum of squares of elements of the vector *x*.

The third model selection method is the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002). The DIC is defined as

$$\text{DIC}=\overline{D(\theta)}+p_D=-4E_{\theta_{\mathbf{k}}}[\log p(\boldsymbol{y}|\theta_{\mathbf{k}})|\boldsymbol{y}]+2\log p(\boldsymbol{y}|\bar{\theta}_{\mathbf{k}}),$$

where $D(\boldsymbol{\theta}_k) = -2 \log p(\mathbf{y}|\theta_k)$ is the deviance and $\overline{D(\theta_k)}$ is the average posterior deviance, $p_D=\overline{D(\theta_k)} - D(\bar{\theta}_k)$ is the "effective dimension", and $\bar{\theta}_k$ is an estimate of $\boldsymbol{\theta}_k$ based on the data $\boldsymbol{y}$. The posterior mean $E[\boldsymbol{\theta}_k|y]$ is often a popular choice for $\bar{\theta}_k$ and is the choice that is implemented in the popular WinBUGS (2005) software, but other choices such as posterior median or mode can also be used. Recently, Celeux et al. (2006) have pointed out that the "effective dimension" $p_D$ can, in fact, be negative in the case of a mixture of distributions. For mixture models, Celeux et al. (2006) suggested eight different modifications of the DIC. The semiparametric model we proposed here utilizes a mixture structure, and we choose $\text{DIC}_3$ (based on terminology used in Celeux et al. (2006)) defined

$$\text{DIC}_3=-4E_{\theta_k}[\log p(\boldsymbol{y}|\theta_{\mathbf{k}})|\boldsymbol{y}]+2\log E_{\theta_{\mathbf{k}}}[p(\boldsymbol{y}|\theta_{\mathbf{k}})|\boldsymbol{y}].$$

Note that the second term is simply based on predictive distribution $p(\mathbf{y}|y) = E_{\boldsymbol{\theta}_k}[p(\boldsymbol{y}|\boldsymbol{\theta}_{\mathbf{k}})|\boldsymbol{y}]$, and the model with the smallest value of $\text{DIC}_3$ will be selected.

## 5. Application

We apply our proposed models to evaluate the survival trend on distant stage testicular cancer survival data obtained from the SEER program. To avoid the accurate specification of cause of death, in the following, we conduct only a relative survival analysis based on (4).

We analyze distant testicular cancer cases diagnosed from 1975 to 2003 with follow-up to 2004, from the SEER 9 registries. Thus, the maximum follow-up time is 29 years. There are 2039 patients diagnosed with distant testicular cancer included in this data for analysis, which is about 12% of patients diagnosed with all stage testicular cancer in the SEER 9 registries during 1975–2003. The $k$-year actuarial survival probabilities with $k = 1, 3, 5$, and 10, over the year of diagnosis, for these patients are presented in Fig. 1.

We apply the proposed joinpoint model on the testicular data with maximum number of joinpoints $K = 3$. Our primary reason for using a maximum of three joinpoints is that, in practice, cancer survival trends typically do not depict too many changes in the overall trends. In order to find the best model, we use the model selection criteria described in Section 5, and compare between models, starting with no joinpoints, $K = 0$, and moving up to $K = 3$ joinpoints. For each value of $k = 0, 1, 2, 3$ we compare the following four models.

**Model 1:** Model with discrete prior for the joinpoints and normal distribution for the random effect $\delta_k$.

**Model 2:** Model with continuous prior for the joinpoints and normal distribution for the random effect $\delta_k$.

**Model 3:** Model with discrete prior for the joinpoints and DPM prior for the random effect $\delta_k$.

**Model 4:** Model with continuous prior for the joinpoints and DPM prior for the random effect $\delta_k$.

We assign weakly informative priors so that the analysis is dominated by the likelihood. Specifically, we assume $\beta \sim N(0, 1000)$. We do not use covariate $z$ for this analysis. For the concentration parameter $\nu$ of the DP prior we assume a $\nu \sim$ Gamma(.1, .1) distribution. This choice of $\nu$ has a prior mean of 1. Note that $\nu = 1$ signifies that the probability of generating a new cluster is $\frac{1}{J+1}$ when we have a sample of size $J$. To assess the effect of this parameter on the inferences, we also considered a Gamma(2, 0.1) distribution for the concentration parameter, and found the results to be very similar. We assume $\eta^{-2} \sim IG(0.1, 0.1)$ and $\pi_k \sim$ Beta(1, 1) for the rest of the DP parameter.

For the continuous prior of the joinpoints, the following characterization of the Dirichlet distribution, in terms of gamma random variables, is used to facilitate the computation. Let $g_u = \frac{(\tau_u - \tau_{u-1})}{(x_m - x_1)}$, where $\tau_0 = x_1$ and $\tau_{K+1} = x_m$. Let $g_1, g_2, \ldots, g_{K+1}$ be independent with $g_u \sim$ Gamma($a_u$, 1). Then $g_{u'} / \sum_u^{K+1} g_u$, $\underline{u}' = 1,2, \ldots, K$ follows a *Dirichlet* ($a_1, \ldots, a_{K+1}$); see Willks (1962). We assume a Gamma(1, 1) distribution prior for $g_u$.

The posterior distributions are analytically intractable. We use a Gibbs sampler (Gelfand and Smith, 1990) to obtain samples from the posterior distributions. We implement our model in the publicly available software `WinBUGS` and `R` (R Development Core Team, 2006). We ran two chains of the Gibbs sampler with widely dispersed initial values. The initial values for the fixed parameters were selected by starting with the prior mean and covering ±3 standard deviations. The initial values for the precision were arbitrarily selected. For each model parameter, the posterior distribution was examined visually by monitoring the sample traces, means and density estimates from the two sequences, as well as by observing the corresponding *R*-statistic (Gelman and Rubin, 1992). Each sequence was run for 25,000 iterations with a burn in of 10,000 samples. Thus, the remaining 15,000 samples in each sequence were combined to yield a total of 30,000 samples upon which the posterior inference is based.

Table 1 presents the comparison among models with different numbers of joinpoints using the three different model selection criteria. Note that higher LPML values and lower EPD and lower $DIC_3$ values correspond to models with better fit, and we do not have the model selection values under DP when $K = 0$ or $K = 1$ as in this case we have zero and only one random slope, respectively. We note that the model with $k = 2$ joinpoints always has higher LPML values and lower $DIC_3$ and EPD values. Here, the measures for models with the DPM prior are consistently lower than those with normal priors for the slope $\delta_k$ when the priors for the joinpoints are the same (model 3 vs. model 1 and model 4 vs. model 2). Thus the DPM prior gives a better fit. There is no advantage in using the continuous prior for joinpoints over the discrete prior in this case. Even though continuous priors for joinpoints bring more accurate estimates for the joinpoints, it also brings more penalty because of the increasing parameter dimension. Actually, for this testicular analysis, the DPM model with a discrete joinpoint prior has the best numerical values.

From the model selection results in Table 1, we note that Model 3 with two joinpoints is the best fitting model. Thus, we report the parameter estimates for $k = 2$ under Model 3 in Table 2.

The point estimates of the two joinpoints for the testicular cancer data under Model 3 came out to be at 1977 and 1995. The APC of death rate is decreasing about 30% each year before 1977 with the introduction of platinum therapy in 1974 (Higby et al., 1974) and the establishment of the standard regimen of PVB (cisplatin, vinblastine and bleomycin) in 1977

(Feuer et al., 1994). The improvement in survival is slower but still with an APC of −4% during the period 1977–1995. The improvement stopped around 1995 and the APC of death rate started to increase slightly (estimated slope as 1.8% with credible interval bigger than 1) every year after 1995.

Age may play an important role in survival trend. The average age of diagnosis of the cancer patients changes over time as shown in Fig. 2. We reanalyze the data with age as a covariate. In this case the model with two joinpoints also got selected. However, the estimates of the joinpoint shifted across the models (see Table 2). The coefficient for age is positive, which implies that the death rate increases as age increases. The first joinpoint without age adjustment is around 1977, but it is around 1988 after we included age as a covariate. The second joinpoint, detected at 1995 without age as a covariate, no longer exists. However, another joinpoint around 2000 is detected when we include age as a covariate. The first joinpoint detected by the model around 1988 with age as a covariate may reflect a large treatment improvement for testicular cancer before the mid-1980s, without confounding of age. According to the results from the model with age as a covariate, after 1988, the hazard rate decreases (with negative APC values) for distant stage testicular cancer. We also noticed that the death rate increases by 1.8% after 1995 from the model without age adjustments, but after we include age in the model, the rate becomes flat after the 1990s, which indicates that the pattern of decrease in survival from the model without age may be explained by the increase of age after the early 1990s (Fig. 2). A comparison of the results from models with and without age as a covariate show that it is important to consider age in survival trend analysis.

## 6. Conclusion and discussion

We have proposed parametric and semiparametric Bayesian joinpoint survival models for analyzing grouped survival data and used these models to analyze the survival trend of testicular cancer. Our method can easily be extended to conduct cause-specific survival analysis if the information on cause of death is available and of good quality. Although the semiparametric model gives similar posterior estimates of the parameters compared with parametric model, the semiparametric model gives a better model fit with shorter credible interval for the joinpoints. Three different model selection methods, namely, the LPML, EPD, and the DIC, were adapted and employed to assess model fit, and the performance of the three criteria is very consistent in selecting the best model. We also consider model selection in the model space search. The proposed joinpoint survival model was further extended to include age as a covariate in the testicular cancer example. We have shown that covariates may affect the number of joinpoints.

## Acknowledgments

The views expressed by the last author do not necessarily reflect those of the FDA and NCI. The research of Dr. Ghosh was supported in part by National Institutes of Health (NIH) contract 263-MQ-610994. Dr. Yu was supported in part by the Intramural Research Program of the National Institute on Aging.

## References

Antoniak CE. Mixtures of Dirichlet processes with applications to nonparametric problems. Annals of Statistics. 1974; 2:1152–1174.

Basu, D.; Tiwari, RC. A note on the Dirichlet process. In: Kallianpur, G.; Krishnaih, PR.; Ghosh, JK., editors. Statistics and Probability. In: Essays in Honor of C.R. Rao. North-Holland Publishing; 1982. p. 89-103.

Celeux G, Forbes F, Robert CP, Titterington DM. Deviance information criteria for missing data models. Bayesian Analysis. 2006; 1:651–674.

Carlin BP, Gelfand AE, Smith AFM. Hierarchical Bayesian analysis of changepoint problems. Applied Statistics. 1992; 41:389–408.

Chang YP, Huang WT. Inferences for the linear error-in-variables with changepoint models. Journal of the American Statistical Association. 1997; 92:171–178.

Chen, Ming-Hui; Qi-man, Shao; Joseph George, Ibrahim. Monte Carlo methods in Bayesian computation. Berlin, New York: Springer-Verlag Inc.; 2000.

Cronin KA, Feuer EJ. Cumulative cause-specific mortality for cancer patients in the presence of other causes: A crude analogue of relative survival. Statistics in Medicine. 2000; 19:1729–1740. [PubMed: 10861774]

Denison DGT, Mallick BK, Smith AFM. Automatic Bayesian curve fitting. Journal of the Royal Statistical Society, Series B. 1998; 60:330–350.

Edere F, Axtell LM, Cutler SJ. The relative survival rate: A statistical methodology. National Cancer Institute Monograph. 1961; 6:101–121. [PubMed: 13889176]

Escober MD. Estimating normal means with a Dirichlet process prior. Journal of the American Statistical Association. 1994; 89:268–277.

Escober MD, West M. Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association. 1995; 90:577–580.

Feder PI. The log likelihood ratio in segmented regression. Annals of Statistics. 1975; 3:84–97.

Ferguson TS. A Bayesian analysis of some nonparametric problems. Annals of Statistics. 1973; 1:209–230.

Feuer EJ, Frey CM, Brawley OW. After a treatment breakthrough: A comparison of trial and population-based data for advanced testicular cancer. Journal of Clinical Oncology. 1994; 12:368–377. [PubMed: 7509384]

Feuer EJ, Kessler LG, Baker SG, Triolo HE, Green DT. The impact of breakthrough clinical trials on survival in population based tumor registries. Journal of Clinical Epidemiology. 1991; 44:141–153. [PubMed: 1704907]

Frome EL. The analysis of rates using Poisson regression models. Biometrics. 1983; 39:665–674. [PubMed: 6652201]

Gail M. A review and critique of some models used in competing risk analysis. Biometrics. 1975; 31:209–222. [PubMed: 1164533]

Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association. 1990; 85:398–409.

Gelfand, AE.; Dey, DK.; Chang, H. Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In: Bernardo, JM.; Berger, JO.; Dawid, AP.; Smith, AFM., editors. Bayesian Statistics. Vol. Volume 4. Oxford: Oxford University Press; 1992. p. 147-159.

Gelfand AE, Ghosh SK. Model choice: A minimum posterior predictive loss approach. Biometrika. 1998; 85:1–11.

Gelman A, Rubin D. Inference from alternative simulation using multiple sequences. Statistical Science. 1992:457–472.

Ghosh P, Basu S, Tiwari RC. Bayesian analysis of cancer rates from SEER program using parametric and semiparametric joinpoint regression models. Journal of the American Statistical Association. 2009; 104(486):439–452.

Ghosh P, Rosner G. A semiparametric Bayesian approach to average bioequivalence. Statistics in Medicine. 2007; 26:1224–1236. [PubMed: 16832840]

Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika. 1995; 82:711–732.

Higby DJ, Wallace J, Albert D, Holland JF. Diamminodicholorplatinum in the chemotherapy of testicular tumors. Journal of Urology. 1974; 112:100–104. [PubMed: 4858106]

Ishwaran H, James L. Gibbs sampling methods for stick-breaking priors. Journal of the American Statistical Association. 2001; 96:161–173.

Kass RE, Ventura V, Cai C. Statistical smoothing of neuronal data, Network. Computation in Neural Systems. 2003; 14:5–15.

Kim HJ, Fay M, Feuer EJ, Midthune DN. Permutation tests for regression with application to cancer rates. Statistics in Medicine. 2000; 19:335–351. [PubMed: 10649300]

Laud PW, Ibrahim JG. Predictive specification of prior model probabilities in variable selection. Biometrika. 1996; 83:267–274.

Lerman PM. Fitting segmented regression models by grid search. Applied Statistics. 1980; 29:77–84.

Luo X, Turnbull BW, Clark LC. Likelihood ratio tests for a changepoint with survival data. Biometrika. 1997; 73:85–89.

MacEachern SN. Estimating normal means with a conjugate style Dirichlet process prior. Communications in Statistics: Simulation and Computation. 1994; 23:727–741.

National Center for Health Statistics. Vol. vol. 52. Hyattsville,Maryland: 2003. National Vital Statistics Reports.

Ohlssen D, Sharples LD, Spiegelhalter DJ. Flexible random-effects models using Bayesian semi-parametric models: Applications in institutional comparisons. Statistics in Medicine. 2007; 26:2088–2112. [PubMed: 16906554]

R Development Core Team. Vienna, Austria: R Foundation for Statistical Computing; 2006. R: A language and environment for statistical computing. ISBN: 3-900051-07-0. URL: http://www.R-project.org

Raftery AE, Akman VE. Bayesian analysis of a Poisson process with a change-point. Biometrika. 1986; 73:85–89.

Ries, LAG.; Harkins, D.; Krapcho, M.; Mariotto, A.; Miller, BA.; Feuer, EJ.; Clegg, L.; Eisner, MP.; Horner, MJ.; Howlader, N.; Hayat, M.; Hankey, BF.; Edwards, BK., editors. Bethesda, MD: National Cancer Institute; 2006. SEER Cancer Statistics Review, 1975–2003. http://seer.cancer.gov/csr/1975-2003/. Based on November 2005 SEER data submission, posted to the SEER web site 2006

Sethuraman J. A constructive definition of Dirichlet priors. Statistica Sinica. 1994; 4:639–650.

Sethuraman, J.; Tiwari, RC. Convergence of Dirichlet measure and the interpretation of their parameters. In: Gupta, S.; Berger, JO., editors. Statistical Decisions Theory and Related Topics III. Vol. vol. 2. Academic Press; 1982. p. 305-315.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society, B. 2002; 64:583–639.

Spiegelhalter, D.; Thomas, A.; Best, N.; Lunn, D. WinBUGS User Manual, Version 1.4. MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology & Public Health, Imperial College School of Medicine; 2005. available at: http://www.mrc-bsu.cam.ac.uk/bugs

Tiwari RC, Cronin KC, Davis W, Feuer EJ, Yu B, Chib S. Bayesian model selection for join point regression with application to age-adjusted cancer rates. Applied Statistics. 2005; 54:919–939.

Willks, SS. Mathematical Statistics. New York: Wiley; 1962.

Zelterman D, Grambsch PM, Le CT, Ma ZF, Curtsinger JW. Piecewise exponential curves with smooth transitions. Mathematical Biosciences. 1994; 120:233–250. [PubMed: 8204986]

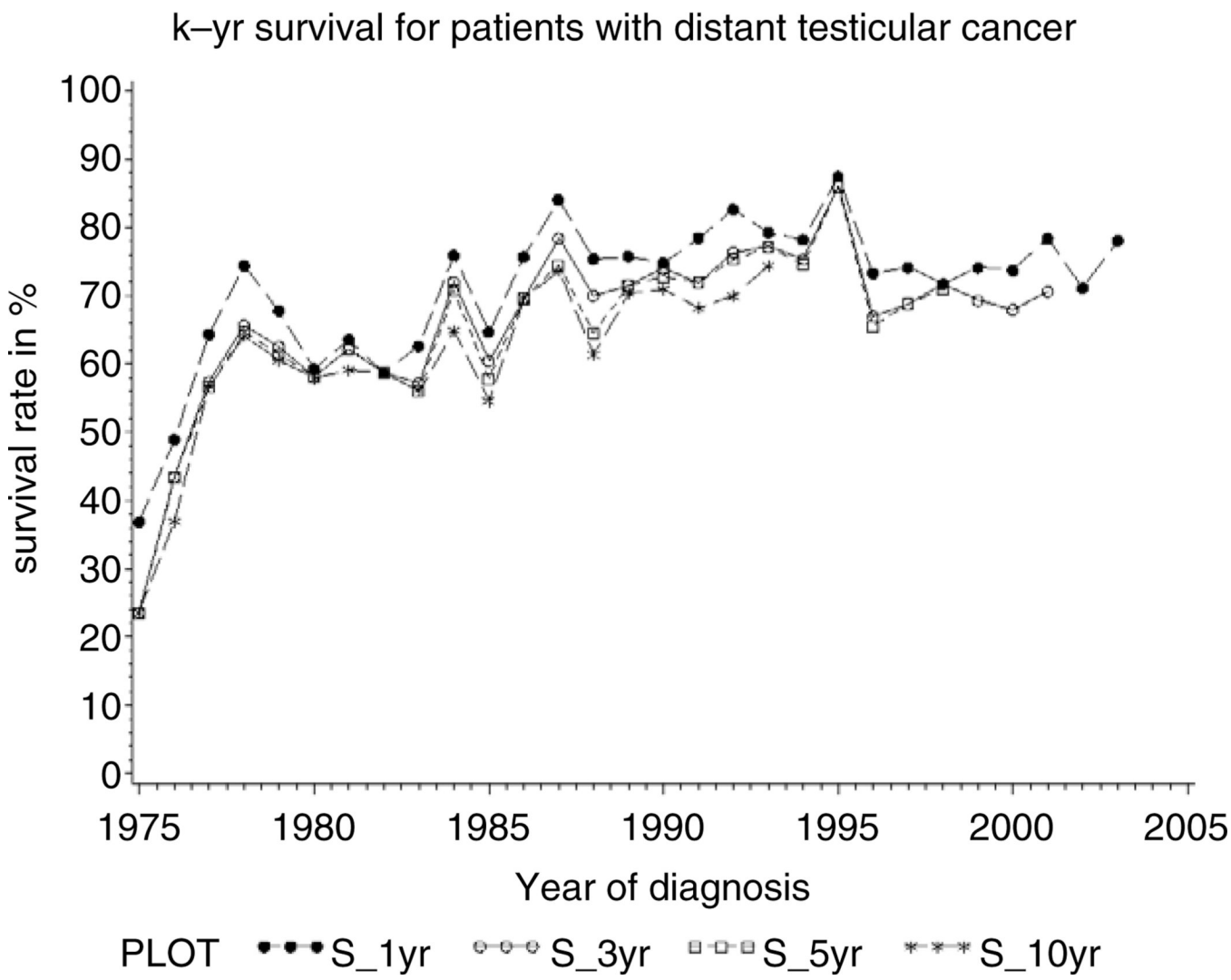


**Fig. 1.**
Actuarial 1-year, 3-year, 5-year, and 10-year cumulative survival probability for patients diagnosed with distant testicular cancer during 1975–2004.

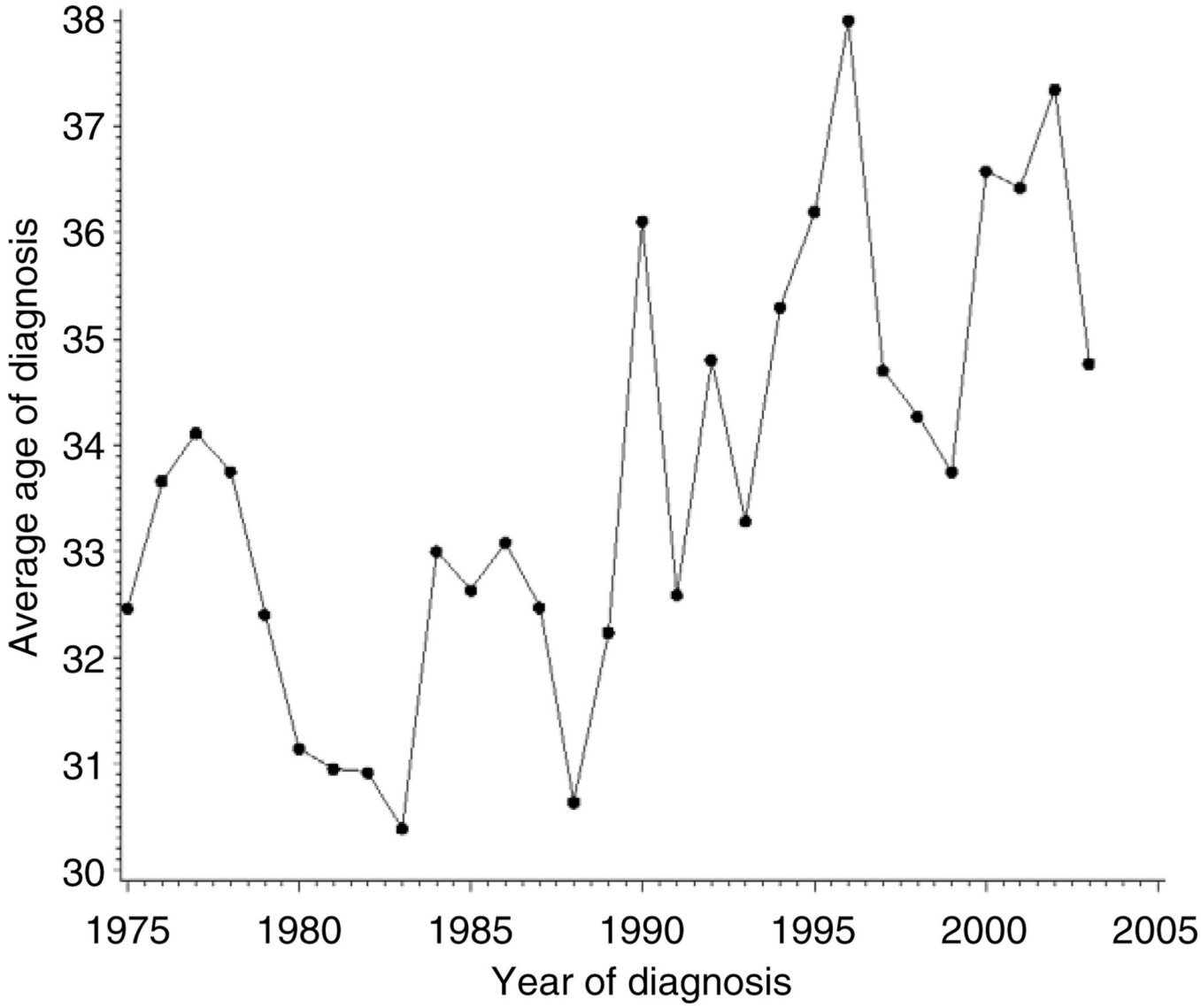


**Fig. 2.**
Average age of diagnosis for patients with distant testicular cancer.

**Table 1**

Model comparison for distant testicular cancer survival data.

| **Model selection criteria** | **Number of joinpoints** | | | |
|---|---|---|---|---|
| | ***K* = 0** | ***K* = 1** | ***K* = 2** | ***K* = 3** |
| | Model 1 | | | |
| LPML | −500.17 | −510.17 | **−489.536** | −506.086 |
| EPD | 58.25 | 51.13 | **48.41** | 52.81 |
| $DIC_3$ | 992.38 | 972.06 | **939.129** | 964.33 |
| | Model 2 | | | |
| LPML | −500.17 | −562.17 | **−490.45** | −585.67 |
| EPD | 58.25 | 52.12 | **51.89** | 53.19 |
| $DIC_3$ | 992.38 | 971.69 | **960.21** | 969.08 |
| | Model 3 | | | |
| LPML | NA | NA | **−484.046** | −489.573 |
| EPD | NA | NA | **48.18** | 51.09 |
| $DIC_3$ | NA | NA | **923.429** | 953.2 |
| | Model 4 | | | |
| LPML | NA | NA | **−488.188** | −583.503 |
| EPD | NA | NA | **48.37** | 51.97 |
| $DIC_3$ | NA | NA | **949.658** | 960.746 |

**Table 2**

Parameter estimates and credible intervals (CIs) under Model 3 for testicular cancer data.

| | Without covariate | With covariate |
|---|---|---|
| $\beta$ | −0.3491 (−0.4743, −0.1626) | −0.46 (−1.7, 0.04) |
| $\gamma$ (age) | – | 0.23 (0.14, 0.43) |
| $\tau_1$ | 1977 (76, 80) | 1988 (83, 94) |
| $\tau_2$ | 1995 (89, 99) | 1999 (90, 02) |
| $APC_1$ | −0.2924 (−0.3777, −0.15) | −0.6309 (−0.95, −0.24) |
| $APC_2$ | −0.0354 (−0.0578, −0.0141) | −0.009 (−0.02, 0.006) |
| $APC_3$ | 0.018 (0.0069, 1.451) | −0.006 (−0.03, 0.009) |