Journal of ***, 2020, *,****
Published Online **** 2020 in JPRM.
http://www.scitecresearch.com/journals****

Change-Point Analysis of Survival Data with Application in Clinical Trials

Xuan Chen[1], **Michael Baron**[2]

[1]Department of Biostatistics and Programming, Sanofi, Beijing, China
Email: Xuan.Chen@sanofi.com

[2]Department of Mathematics and Statistics, American University, Washington DC, USA
Department of Mathematical Sciences, University of Texas at Dallas, Richardson TX, USA
Email: baron@american.edu, mbaron@utdallas.edu

Received **** 2020

Abstract

Effects of many medical procedures appear after a time lag, when a significant change occurs in subjects' failure rate. This paper focuses on the detection and estimation of such changes which is important for the evaluation and comparison of treatments and prediction of their effects. Unlike the classical change-point model, measurements may still be identically distributed, and the change point is a parameter of their common survival function. Some of the classical change-point detection techniques can still be used but the results are different. Contrary to the classical model, the maximum likelihood estimator of a change point appears consistent, even in presence of nuisance parameters. However, a more efficient procedure can be derived from Kaplan-Meier estimation of the survival function followed by the least-squares estimation of the change point. Strong consistency of these estimation schemes is proved. The finite-sample properties are examined by a Monte Carlo study. Proposed methods are applied to a recent clinical trial of the treatment program for strong drug dependence.

Keywords

Change-Point Problem; Failure Rate; Kaplan-Meier Estimation; Least Squares Estimation; Maximum Likelihood Estimation; Strong Consistency; Survival Function

1. Introduction

Change-point models studied in clinical research usually refer to changes in the failure rate. Many articles and clinical reports describe situations when after a certain survival period, the failure rate is expected to change due to the treatment or during the after-treatment recovery. Detection of such changes, their

How to cite this paper: X. Chen and M. Baron (2015) Change-Point Analysis of Survival Data with Application in Clinical Trials. **********,*,***-***

estimation, and their comparison between different groups of patients (the treatment arm and the placebo arm is the classical example) are important understanding the treatment's effect and for the evaluation of the treatment's success. Survival times in this example have a higher initial failure rate and a lower failure rate afterwards. Similar examples are found in [3, 4, 5, 6, 7, 8, 9].

This situation is conceptually and mathematically different from the classical change-point model, see e.g. [10, 11, 12, 13, 14], where observations follow one distribution before the change point and another distribution after it. In the described scenario, with one or several changes in the failure rate, all the subjects are assumed to have the same distribution. Each change point is understood as a parameter of this distribution that separates two patterns, two different models for the failure rate, and typically, it is the moment of a "clinically significant" reduction of the failure rate.

2. Survival models with change points

We assume a constant failure rate function $\lambda(x) = \lambda_0$ until an unknown time τ. Change occurs at time τ, and $\lambda(x)$ shifts to a new value λ_1 and remains at it thereafter. Thus,

$$\lambda(x) = \lambda_0 1_{x\le\tau} + \lambda_1 1_{x>\tau} \tag{1}$$

where $\lambda_0 > 0, \lambda_1 > 0$, and τ is *the change point*, the main parameter of interest.

3. Maximum likelihood estimation

Under model (1), the likelihood function of $X_1, ..., X_n$ is

$$\begin{aligned} F(x_1, ..., x_n|\lambda_0, \lambda_1, \tau) &= \lambda(x) \exp\left\{-\int_0^x \lambda(t)dt\right\} \\ &= \prod_{\delta_i=1} (\lambda_0 e^{-\lambda_0 x_i} 1_{x_i\le\tau} + \lambda_1 e^{-\lambda_0\tau-\lambda_1(x_i-\tau)} 1_{x_i>\tau}) \times \prod_{\delta_i=0} (e^{-\lambda_0 x_i} 1_{x_i\le\tau} + e^{-\lambda_0\tau-\lambda_1(x_i-\tau)} 1_{x_i>\tau}), \end{aligned}$$

which yields the log-likelihood ratio

$$\Lambda(\tau) = \sum_{i\le n} \left\{ \delta_i \left(\log\frac{\lambda_1}{\lambda_0} - (x_i-\tau)(\lambda_1-\lambda_0) \right) 1_{x_i>\tau} + (1-\delta_i)(x_i-\tau)(\lambda_0-\lambda_1) 1_{x_i>\tau} \right\} = \sum_{i\le n} y_i \tag{2}$$

where

$$y_i = \begin{cases} \log\frac{\lambda_1}{\lambda_0} + (\lambda_0-\lambda_1)(x_i-\tau) & \text{for } \tau < x_i \text{ and } \delta_i = 1 \\ (\lambda_0-\lambda_1)(x_i-\tau) & \text{for } \tau < x_i \text{ and } \delta_i = 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

4. Least squares method based on Kaplan-Meier estimation

In this Section, we introduce a different change-point estimation procedure which is based on Kaplan-Meier estimator of the survival function. Since the Kaplan-Meier method is nonparametric, the change-point estimation scheme proposed here can be easily extended to a wide variety of survival models with change points arising in clinical trials and other applications.

Kaplan and Meier (1958) proposed a famous estimator for the survival function $S(t)$:

$$\tilde{S}_n(x) = \prod_{x_{(j)}\le x} \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}} \tag{4}$$

This is a step function with jumps at observations X_i for which $\delta_i = 1$. It is a nonparametric estimator of the survival function, and it can be applied in presence of censoring. No assumptions are required for the

probability distribution other than the independence between the survival and censoring variables. Kaplan-Meier estimator (4) has the following properties:

1. It is the nonparametric maximum likelihood estimator of the true survival function $S(x)$.
2. It has an asymptotically normal distribution for any x where $S(x)$ is continuous.
3. It converges almost surely to $S(x)$ uniformly in x, and for each $\epsilon > 0$, there exists $c > 0$, such that $\mathbf{P}(|\tilde{S}_n(x) - S(x)| > \epsilon) \leq e^{-nc}$ for sufficiently large n. Refer to [20] for details.
4. If no censoring occurs or all variables are censored at the same time, then the Kaplan-Meier estimator reduces to the usual empirical distribution function.

4.1. Least squares estimation and strong consistency

Under the piecewise constant failure rate model (1) with a change point τ, the logarithm of the survival function at the time x_i is given as

$$L_i(\tau, \lambda_0, \lambda_1) = \log S(x_i) = -\lambda_0 x_i 1_{x_i \leq \tau} - (\lambda_0 \tau + \lambda_1(x_i - \tau))1_{x_i > \tau}$$

Let $\theta = (\lambda_0, \lambda_1, \tau)$ denote the vector of parameters. Its least squares estimator $\tilde{\theta} = (\tilde{\tau},\ \tilde{\lambda}_0,\ \tilde{\lambda}_1)$ consists of those values of τ, λ_0, and λ_1 that minimize the error sum of squares

$$\mathrm{ESS}(\theta) = \sum_{i=1}^{n} (\tilde{y}_n(x_i) - L_i(\theta))^2, \tag{5}$$

where

$$\tilde{y}_n(x_i) = \log \tilde{S}_n(x_i) = \sum_{x_{(j)} \leq x_i} \delta_{(j)} \log \frac{n-j}{n-j+1} \tag{6}$$

Lemma 1. *At $\theta = \tilde{\theta}$, the error sum of squares components satisfy the strong law of large numbers; that is, $\frac{1}{n} ESS(\tilde{\theta})$ converges to 0 almost surely, as $n \to 0$.*

The proof can be found in the Appendix.

To prove the strong consistency of the vector of least squares estimators $\tilde{\tau},\ \tilde{\lambda}_0,\ \tilde{\lambda}_1$, we express $\frac{1}{n}\mathrm{ESS}(\tilde{\theta})$ in terms of the residual $\alpha_n(x)$,

$$\begin{aligned}
\mathrm{ESS}(\tilde{\theta}) &= \sum_{x_i \leq \tilde{\tau}} (\tilde{y}_n(x_i) + \tilde{\lambda}_0 x_i)^2 + \sum_{x_i > \tilde{\tau}} (\tilde{y}_n(x_i) + \tilde{\lambda}_0 \tilde{\tau} + \tilde{\lambda}_1(x_i - \tilde{\tau}))^2 \\
&= \sum_{x_i \leq \min(\tilde{\tau}, \tau)} (-\lambda_0 x_i + \tilde{\lambda}_0 x_i + \alpha_n(x_i))^2 + \sum_{x_i \geq \max(\tilde{\tau}, \tau)} (-\lambda_0 \tau - \lambda_1(x_i - \tau) + \tilde{\lambda}_0 \tilde{\tau} + \tilde{\lambda}_1(x_i - \tilde{\tau}) + \alpha_n(x_i))^2 \\
&\quad + \sum_{\tilde{\tau} < x_i \leq \tau} (-\lambda_o x_i + \tilde{\lambda}_0 \tilde{\tau} + \tilde{\lambda}_1(x_i - \tilde{\tau}) + \alpha_n(x_i))^2 + \sum_{\tau < x_i \leq \tilde{\tau}} (-\lambda_o \tau - \lambda_1(x_i - \tau) + \tilde{\lambda}_0 x_i + \alpha_n(x_i))^2. \\
&= A_n + B_n + C_n + D_n,
\end{aligned} \tag{7}$$

where

$$\begin{aligned}
A_n &= \sum_{X_i \leq \min(\tilde{\tau}, \tau)} (-\lambda_0 x_i + \tilde{\lambda}_0 x_i)^2, \\
B_n &= \sum_{X_i \geq \max(\tilde{\tau}, \tau)} (-\lambda_0 \tau - \lambda_1(x_i - \tau) + \tilde{\lambda}_0 \tilde{\tau} + \tilde{\lambda}_1(x_i - \tilde{\tau}))^2, \\
C_n &= \sum_{\tilde{\tau} < X_i \leq \tau} (-\lambda_0 x_i + \tilde{\lambda}_0 \tilde{\tau} + \tilde{\lambda}_1(x_i - \tilde{\tau}))^2, \\
D_n &= \sum_{\tau < X_i \leq \tilde{\tau}} (-\lambda_0 \tau - \lambda_1(x_i - \tau) + \tilde{\lambda}_0 x_i))^2.
\end{aligned}$$

The uniform convergence of $\alpha_n(x)$ and the strong law of large numbers in [21] imply directly that

$$A_n/n \xrightarrow{a.s.} 0, \tag{8}$$
$$B_n/n \xrightarrow{a.s.} 0, \tag{9}$$
$$C_n/n \xrightarrow{a.s.} 0, \tag{10}$$
$$D_n/n \xrightarrow{a.s.} 0. \tag{11}$$

Since we assume that there is indeed a change-point, it is reasonable to make the following assumption.
Assumption (A): There exist known $0 < m < M$ such that $\tau_0 \in [m, M]$.

Theorem 1. *$\tilde{\lambda}_0$ is strongly consistent for λ_0 under Assumption (A).*

The proof can be found in the Appendix.

Theorem 2. *$\tilde{\tau}$ is strongly consistent for τ under Assumption (A).*

Proof. (i) We will prove $\mathbf{P}(\tilde{\tau} \to \tau \cap \tilde{\tau} > \tau) = 1$ in this part.
We prove by contradiction. Suppose for any $\epsilon > 0$, there exist $\delta > 0$ and $N(\epsilon)$ such that

$$\mathbf{P}\left(\tilde{\tau} - \tau > \epsilon \cap \tilde{\tau} > \tau\right) > \delta \quad \text{for all} \quad n > N(\epsilon). \tag{12}$$

From Theorem 1 and (11), we get

$$\frac{1}{n} \sum_{\tau < X_i \le \tilde{\tau}} (x_i - \tau)^2 (\lambda_0 - \lambda_1)^2 \xrightarrow{a.s.} 0 \tag{13}$$

From (12), we have

$$\mathbf{P}\left(\sum_{\tau < X_i \le \tilde{\tau}} (x_i - \tau)^2 > \sum_{0 < X_i - \tau \le \epsilon} (x_i - \tau)^2 \right) > \delta$$

for all $n > N(\epsilon)$.
Also,

$$\frac{1}{n} \sum_{0 < X_i - \tau \le \epsilon} (x_i - \tau)^2 \xrightarrow{a.s.} \mathbf{E}(X - \tau)^2 \mathbf{1}_{0 < X - \tau \le \epsilon} > 0.$$

Hence, for sufficiently large n,

$$\mathbf{P}\left(\frac{1}{n} \sum_{\tau < X_i \le \tilde{\tau}} (x_i - \tau)^2 > \frac{1}{n} \sum_{0 < X_i - \tau \le \epsilon} (x_i - \tau)^2 > \frac{1}{2} \mathbf{E}(X - \tau)^2 \mathbf{1}_{0 < X - \tau \le \epsilon} > 0 \right) > \delta/2,$$

which contradicts (13).

(ii) We will prove $\mathbf{P}(\tilde{\tau} \to \tau \cap \tilde{\tau} \le \tau) = 1$ in this part.
We also prove this by contradiction. Suppose for any $\epsilon > 0$, there exist $\delta > 0$ and $N(\epsilon)$ such that

$$\mathbf{P}\left(\tilde{\tau} - \tau > \epsilon \cap \tilde{\tau} \le \tau\right) > \delta \tag{14}$$

for all $n > N(\epsilon)$.
Then

$$\mathbf{P}\left(\frac{1}{n} \sum_{\tilde{\tau} < X_i \le \tau} (x_i - \tilde{\tau})^2 > \frac{1}{n} \sum_{\tau - \epsilon < X_i \le \tau} (x_i - \tau + \epsilon)^2 \right) > \delta$$

for all $n > N(\epsilon)$.

Also,

$$\frac{1}{n}\sum_{\tau-\epsilon<X_i\leq\tau}(x_i-\tau+\epsilon)^2 \xrightarrow{a.s.} \mathbf{E}(X-\tau+\epsilon)^2\mathbf{1}_{\tau-\epsilon<X\leq\tau}>0.$$

Hence,

$$\mathbf{P}\left(\frac{1}{n}\sum_{\tilde{\tau}<X_i\leq\tau}(x_i-\tilde{\tau})^2>0\right)>\delta \quad \text{for} \quad n\to\infty$$

From (10) and Theorem 1, we can get

$$\mathbf{P}(\tilde{\lambda}_1\to\lambda_0)>\delta \quad \text{for} \quad n\to\infty$$

Hence

$$\mathbf{P}\left(\frac{1}{n}\mathrm{ESS}(\tilde{\theta})-\frac{1}{n}\sum_{i=1}^{n}(-\lambda_0\tau-\lambda_1(x_i-\tau)+\lambda_0 x_i)^2\to 0\right)>\delta,$$

whereas

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(-\lambda_0\tau-\lambda_1(x_i-\tau)+\lambda_0 x_i)^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\tau)^2(\lambda_0+\lambda_1)^2\\ &\xrightarrow{a.s.} (\lambda_0+\lambda_1)^2\mathbf{E}(x-\tau)^2\mathbf{1}_{X>\tau}>0.\end{aligned}$$

Hence

$$\mathbf{P}\left(\frac{1}{n}\mathrm{ESS}(\tilde{\theta})>\frac{1}{2n}\sum(-\lambda_0\tau-\lambda_1(x_i-\tau)+\lambda_0 x_i)^2>0\right)>\delta/2$$

for sufficiently large n, which contradicts Theorem 1.

Combining (i) and (ii) gives

$$\mathbf{P}(\tilde{\tau}\to\tau)=1.$$

□

Theorem 3. *$\tilde{\lambda}_1$ is strongly consistent for λ_1 under Assumption (A).*

The proof can be found in the Appendix.

4.2. Convergence rate of the least squares estimator

Now let us investigate the convergence rate of $\tilde{\tau}$ for known λ_0 and λ_1. We will analyze the probability that $\mathrm{ESS}(\tau)$ is less than $\mathrm{ESS}(\tau_0)$ for τ outside of the ϵ-neighborhood of τ_0, where τ_0 is the true value of the change point.

Theorem 4. *For any $\epsilon>0$, there exists $c>0$, such that*

$$\mathbf{P}\left(\bigcup_{\tau:|\tau-\tau_0|>\epsilon}\{ESS(\tau)-ESS(\tau_0)<0\}\right)\leq e^{-nc}$$

for sufficiently large n.

The proof can be found in the Appendix.

Corollary 1. *The change-point estimator $\tilde{\tau}$ is strongly consistent; $\tilde{\tau}\to\tau_0$ almost surely as $n\to\infty$. In particular, for any $\epsilon>0$, there exists $c>0$ such that*

$$\mathbf{P}\left(|\tilde{\tau}-\tau_0|>\epsilon\right)\leq e^{-nc}$$

for sufficiently large n.

Proof. According to Theorem 4, for any arbitrary sequence $\epsilon_j > 0$, $\epsilon_j \downarrow 0$ as $j \uparrow \infty$, there exists $c(\epsilon_j) > 0$ such that $P(\alpha_n \geq \epsilon_j) \leq e^{-nc(\epsilon_j)}$. Hence

$$\sum_{n=1}^{\infty} \mathbf{P}\left(\bigcup_{\tau:|\tau-\tau_0|>\epsilon_j} \{\mathrm{ESS}(\tau) - \mathrm{ESS}(\tau_0) < 0\} \right) \leq \sum_{n=1}^{\infty} e^{-nc(\epsilon_j)} = \frac{e^{-c(\epsilon_j)}}{1 - e^{-c(\epsilon_j)}} < \infty$$

Since the sum of probabilities converges, by the Borel-Cantelli lemma, with probability one, $\mathrm{ESS}(\tau) \geq \mathrm{ESS}(\tau_0)$ for all $\tau : |\tau - \tau_0|$ for sufficiently large n. Therefore, $\tilde{\tau}$, the minimizer of $\mathrm{ESS}(\tau)$, belongs to the ϵ_j-neighborhood of τ_0 almost surely and all sufficiently large n.

It remains to let ϵ_j go to zero over a countable set (e.g., $\epsilon_j = 1/j$). For each j, we obtain that $|\tilde{\tau} - \tau_0| \leq \epsilon_j$ almost surely. Therefore, $\tilde{\tau} \to \tau_0$ a.s., as $n \to \infty$.

□

5. Least squares method for the Cox proportional hazard model with a change point

Generalizing the previous results, in this Section we develop change-point estimation techniques for a more general model, *Cox proportional hazard model with a change point*. Under this model, the hazard rate function has the form,

$$h(x|Z) = h_0(x) \exp(\beta_0' Z) \mathbf{1}_{x \leq \tau} + h_1(x) \exp(\beta_1' Z) \mathbf{1}_{x > \tau} \tag{15}$$

where Z is a vector of covariates $(z_1, ..., z_k)$, β_0', β_1' are vectors of coefficients, and $h_0(x), h_1(x)$ are baseline hazard rates. Clearly, a model with covariates allows to study effects of numerical and categorical factors on the occurrence of a change point and to compare change points between subpopulations.

It is well known that Cox proportional hazard model is *semiparametric*. Indeed, it puts no assumptions on the form of baseline hazard rates $h_0(x)$ and $h_1(x)$ (nonparametric part of model) but assumes a parametric form of the effect of covariates on the hazard.

Introduce the following notations:

- $\lambda_0(x|Z) = h_0(x) \exp(\beta_0' Z)$ is the hazard function before the change point;
- $\lambda_1(x|Z) = h_1(x) \exp(\beta_1' Z)$ is the hazard function after the change point;
- $F(x_1, ..., x_n|Z)$ is the joint likelihood function under model (15);
- $\Lambda(\tau|Z)$ is log-likelihood ratio under model (15);
- $S(x|Z)$ is survival function under model (15);
- $\Theta = (\tau, \beta_0, \beta_1)$ is the unknown parameter vector;
- $\tilde{\Theta}$ is the least squares estimator of Θ which, similarly to Section 4.1, minimizes the error sum of squares based on the differences between the log-survival functions obtained from model (15) and from the Kaplan-Meier estimator (4).

Under model (15), the survival function is expressed as

$$S(x|Z) = \exp\left(-\int_0^x \lambda_0(s|Z) ds \right) \mathbf{1}_{x \leq \tau} + \exp\left(-\int_0^\tau \lambda_0(s|Z) ds - \int_\tau^x \lambda_1(s|Z) ds \right) \mathbf{1}_{x > \tau},$$

so that

$$L_i(\tau|Z) = \log S(x_i|Z) = -\int_0^{x_i} \lambda_0(s|Z) ds) \mathbf{1}_{x_i \leq \tau} - \left(\int_0^\tau \lambda_0(s|Z) ds + \int_\tau^{x_i} \lambda_1(s|Z) ds \right) \mathbf{1}_{x_i > \tau}.$$

The least squares estimator $\tilde{\Theta} = (\tilde{\tau}, \tilde{\beta}_0, \tilde{\beta}_1)$ of the change point τ_0 and slopes β_0 and β_1 is then defined as the minimizer

$$\tilde{\Theta} = \arg\min_{\Theta} \mathrm{ESS}(\Theta|Z)$$

of the error sum of squares

$$\mathrm{ESS}(\Theta|Z) = \sum_{i=1}^{n} (\tilde{y}_n(x_i) - L_i(\tau|Z))^2, \tag{16}$$

where components $\tilde{y}_n(x_i)$ are defined in (6).

5.1. Strong consistency and convergence rate of the least squares estimator

Extention of the results of Section 4 on the strong consistency of the change point estimator and estimators of the nuisance parameters to Cox proportional hazard model is straightforward. Indeed, the uniform strong consistency of the Kaplan-Meier estimator holds for any type of the underlying distribution of survival times. Therefore, the error sum of squares can be split into four parts as in (7), with almost sure convergence holding for each part.

Along the same lines as in the constant hazard rate model, we obtain the following results.

Lemma 2. *At $\Theta = \tilde{\Theta}$, components of the error sum of squares (16) satisfy the strong law of large numbers; that is, $\frac{1}{n}ESS(\tilde{\Theta}|Z)$ converges almost surely to 0 as $n \to 0$.*

Theorem 5. *With known β_0 and β_1, the change-point estimator $\tilde{\tau}$ is strongly consistent. It converges to the true change point τ_0 at the same rate as in the constant hazard rate model; i.e., for any $\epsilon > 0$,*

$$\mathbf{P}\left(|\tilde{\tau} - \tau_0| > \epsilon\right) \leq e^{-nc}$$

for some $c > 0$ and all sufficiently large n.

Proof. The proof is similar to the proof of Theorem 4.5 and Corollary 4.6 of Section 4.2. □

The following results show that the strong consistency of $\tilde{\tau}$ holds even without the assumption of known slopes β_0 and β_1.

Theorem 6. *The estimated slopes $\tilde{\beta}_0$ and $\tilde{\beta}_1$ are strongly consistent for β_0 and β_1 under Assumption (A).*

Theorem 7. *Under unknown slope parameters β_0 and β_1, the change-point estimator $\tilde{\tau}$ is strongly consistent under Assumption (A).*

Strong consistency of $\tilde{\tau}$ and $\tilde{\beta}_i$ in presence of nuisance parameters is proved by the techniques developed in Section 4.1 and essentially along the same lines. For details, see [22], chapter 5.

6. Comparison of estimators

In classical cases, under the usual regularity assumptions, the maximum likelihood estimator is asymptotically the uniformly minimum variance unbiased estimator. Change-point models violate the regularity conditions because of the discontinuity of the likelihood function at the change-point parameter. As a result, the maximum likelihood estimator may no longer be optimal.

An example of ESS, a piecewise polynomial function, is depicted in Figure 1.

Table 1 lists the estimates of τ_0, λ_0, and λ_1 for different sample size and different actual failure rates. Table 2 lists the mean square errors for estimates of τ_0, λ_0, and λ_1. These estimates and mean square errors lead to the following conclusions:

1. Both MLE and LSE of τ_0, λ_0, and λ_1 converge to the true change point and hazard rates as the sample size increases.

2. Both MLE and LSE become more accurate when the difference between λ_0 and λ_1 is increased, holding the sample size constant.

Figure 1: Error sum of squares and the least squares estimator of the change-point

λ_0	λ_1	Sample Size	MLE			LSE		
			τ_0	λ_0	λ_1	τ_0	λ_0	λ_1
0.3	0.1	100	2.8	0.33	0.150	3.925	0.239	0.159
		200	2.701	0.315	0.156	5.117	0.233	0.157
		300	2.979	0.312	0.147	5.917	0.222	0.155
0.25	0.15	100	2.809	0.271	0.173	3.860	0.234	0.188
		200	2.93	0.263	0.176	3.808	0.254	0.184
		300	3.146	0.262	0.171	4.232	0.251	0.182
0.2	0.15	100	3.44	0.208	0.161	4.136	0.212	0.169
		200	3.403	0.208	0.159	4.72	0.225	0.166
		300	3.261	0.208	0.158	5.111	0.242	0.164

Table 1: Estimates of τ_0, λ_0, and λ_1 from Simulated Data

3. The LSE of τ_0 has a lower bias than the MLE for the same sample size and the same failure rates. The mean squared error of the LSE of τ_0 is larger than that of the MLE, for the same sample size and same failure rates, however, the hazard rates are estimated by the LSE method with the same or lower mean square error.

7. Example: Prometa clinical trial

In this section, we apply both the maximum likelihood method and the least squares method to a recent clinical trial for treating methamphetamine-dependent patients conducted by Research Across America, an outpatient clinical research center in Dallas, Texas [17].

Fifty patients participated in an open-label study over the time frame of 84 days. In this study, all of the participants were long-term users of methamphetamine. After the screening visit on day 0, patients received five infusions during the first three weeks and conducted 14 follow-up visits.

Later, a double-blind, placebo-controlled study was conducted to better evaluate the effect of treatment. In the double-blind study, neither the participants nor the clinicians knew which patients belong to which treatment arm. The reason for blinding and placebo controls is to determine (as much as possible) whether the effects observed in the study are due to the treatment itself and not other factors. For each participant, the survival time is the time to relapse, which is the duration of time without the use of drugs.

Our goal here is to detect the *after-treatment effect* of Prometa, which results in a significant reduction of failure rate some time *after* the first three infusions. We detect such changes with both the maximum

λ_0	λ_1	Sample Size	MSE for MLE			MSE for LSE		
			τ_0	λ_0	λ_1	τ_0	λ_0	λ_1
0.3	0.1	100	10.005	0.112	0.025	15.919	0.059	0.026
		200	7.98	0.101	0.025	29.864	0.059	0.025
		300	9.7615	0.098	0.022	38.455	0.055	0.024
0.25	0.15	100	10.239	0.076	0.031	16.177	0.057	0.036
		200	9.549	0.07	0.032	20.361	0.077	0.034
		300	11.238	0.069	0.03	25.67	0.071	0.033
0.2	0.15	100	12.609	0.044	0.028	19.848	0.055	0.03
		200	12.799	0.044	0.026	29.5	0.064	0.028
		300	12.161	0.044	0.025	34.978	0.038	0.027

Table 2: Mean Squared Errors of Estimates of τ_0, λ_0, and λ_1

	Open-label Study			Male Group			Female Group		
	τ	λ_0	λ_1	τ	λ_0	λ_1	τ	λ_0	λ_1
MLE	13	0.1402	0.0105	8	0.1649	0.0201	17	0.1387	0
LSE	14.2	0.1281	0.0142	14	0.1494	0	13	0.1495	0

Table 3: Estimates of $\tau, \lambda_0, \lambda_1$ for Open-label Study

likelihood method and the least squares method. Results are listed in Tables 3 and 4.

First, we estimate the change point for the 50-subject open-label study.

(1) Using the maximum likelihood method, day 13 maximizes the log-likelihood ratio in Figure 2, left. The likelihood ratio test provides a p-value of $1.5067 \cdot 10^{-11}$, which is low enough to reject the null hypothesis "there is no change point". On the day of the change, the failure rate drops from 0.1402 to 0.0105. Thus, we conclude that the failure rate after taking the drugs reduces significantly from 0.1402 to 0.0105 if the patients do not use drugs for 13 days following the treatment.

(2) Using the least squares method, the estimate for change point is 14.2373 and the failure rate drops from 0.1281 to 0.0142, which are very close to the results from maximum likelihood estimate. The graph of error sum of squares is in Figure 2, right.

Change points for the female and male groups are compared to see whether occurrence of a change point depends on gender.

(1) Using the method of maximum likelihood, the estimated change points for males and females are 8 and 17 from Figure 3, left. However, the likelihood ratio test fails to detect a significant difference between the genders with the p-value of 0.3203, i.e., there is no evidence that there are any significantly different change points for males and females. The failure rate reduces from 0.1649 to 0.0201 for males and from 0.1387 to practically 0 for females.

	Prometa Group			Placebo Group		
	τ	λ_0	λ_1	τ	λ_0	λ_1
MLE	13	0.0781	0.0139	18	0.1145	0.0532
LSE	17	0.0720	0	14	0.1255	0.0016

Table 4: Estimates of $\tau, \lambda_0, \lambda_1$ for Two-armed Double-blind Study

Figure 2: Least squares estimate of change-point for open-label study

Figure 3: Least Squares estimate of change-point for female and male groups

(2) Using the least squares method, the change-point estimator for males is about day 14 and the failure rate reduces from 0.1494 to almost 0, while the change-point estimator for females is 13 and the failure rate reduces from 0.1495 to almost 0. We can see that there is almost no difference between male group and female group in change-point estimators from graph 3, right.

Finally, we estimated the change points for the randomized double-blind placebo-controlled study. Change points are estimated separately for the active treatment group and for the placebo group.

(1) The graph of log-likelihood ratios is in Figure 4, left. The estimated change point for the treatment group is 13, and the failure rate reduces from 0.0781 to 0.0139. For the placebo group, the change-point estimate is 18, and the failure rate reduces from 0.1145 to 0.0532. The likelihood ratio test shows that these two groups have significantly different change points with p-value 0.0098.

(2) With the least squares method, the change-point estimator for the treatment group is around day 17 and the failure rate reduces from 0.0720 to almost 0, while the change-point estimator for Placebo is around 14 and the failure rate reduces from 0.1255 to 0.0016. The graph for error sum of squares is in 4, right.

As a result, besides statistical significance, existence of change-points in the survival curves for both treatment groups has important clinical significance. It shows a drop in the risk of relapse after a certain period of abstinence. Although the MLE and LSE methods slightly disagree on the exact location of change-points in the two treatment groups, both methods show that the after-change failure rate is significantly lower for the active treatment groups. Essentially, a patient has to abstain from methamphetamine for two weeks after receiving the treatment, and then the failure rate reduces significantly.

Figure 4: Least Squares estimate of change-point for Prometa and Placebo groups

Acknowledgments

We thank the Editor and the referee for their comments. Research of M. Baron is funded by the National Science Foundation grant DMS 1322353. This support is greatly appreciated.

References

[1] R. A. Desmond, H. L. Weiss, R. B. Arani, S.-J. Soong, M. J. Wood, P. Fiddian, J. Gnann, and R. J. Whitley. Clinical applications for change-point analysis of herpes zoster pain. J. of Pain and Symptom Management, 23(6):510–516, 2002.

[2] D. M. Zucker and E. Lakatos. Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. Biometrika, 77:853–864, 1990.

[3] M. S. Goodman, Y. Li, and R. C. Tiwari. Detecting multiple change points in piecewise constant hazard functions. J. of Applied Statistics, 38(11):2523–2532, 2011.

[4] P. He, G. Kong, and Z. Su. Estimating the survival functions for right-censored and interval-censored data with piecewise constant hazard functions. Contemporary Clinical Trials, 36:122–127, 2013.

[5] C. R. Loader. Inference for a hazard rate change point. Biometrika, 78(4):749–757, 1991.

[6] D. E. Matthews and V. T. Farewell. On testing for a constant hazard against a change-point alternative. Biometrics, 38:463–468, 1982.

[7] H. G. Muller and J. L. Wang. Nonparametric analysis of changes in hazard rates for censored survival data: An alternative to change-point models. Biometrika, 77(2):305–314, 1990.

[8] H. T. Nguyen, G. S. Rogers, and E. A. Walker. Estimation in change-point hazard rate models. Biometrika, 71(2):299–304, 1984.

[9] D. Sertkaya and M. T. Sözer. A Bayesian approach to the constant hazard model with a change point and an application to breast cancer data. Hacettepe J. of Math. and Stat., 32:33–41, 2003.

[10] M. Ahsanullah, A. L. Rukhin, and B. Sinha. Applied Change Point Problems in Statistics. Nova Science Publishers, New York, 1995.

[11] M. Basseville and I. V. Nikiforov. Detection of Abrupt Changes: Theory and Application. PTR Prentice-Hall, Inc., Englewood Cliffs, NJ, 1993.

[12] P. K. Bhattacharya. Some aspects of change-point analysis. In Change-Point Problems, pages 28–56, IMS Lecture Notes - Monograph Series (Volume 23), 1995.

[13] J. Chen and A. K. Gupta. Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance. Birkhäuser, Boston, MA, 2012.

[14] H. V. Poor and O. Hadjiliadis. Quickest Detection. Cambridge University Press, Cambridge, UK, 2009.

[15] K. J. Worsley. Exact percentage points of the likelihood-ratio test for a change-point hazard-rate model. Biometrics, 44(1):259–263, 1988.

[16] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc., 53:457–481, 1958.

[17] H. C. Urschel, L. L. Hanselka, I. Gromov, L. White, and M. Baron. Open-label study of a proprietary treatment program targeting type a γ-aminobutyric acid receptor dysregulation in methamphetamine dependence. Mayo Clinic Proceedings, 82(10):1170–1178, 2007.

[18] Y.-C. Yao. Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. Ann. Statist., 15:1321–1328, 1987.

[19] O. E. Barndorff-Nielsen and D. R. Cox. Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. J. Royal Statist. Soc. B, 46, 1984.

[20] I. H. Dinwoodie. Large deviations for censored data. Ann. Statist., 21:1608–1620, 1993.

[21] P. Billingsley. Probability and Measure, 3rd edition. Wiley, New York, 1995.

[22] X. Chen. Change-point analysis of survival data with application in clinical trials. Ph. D. Dissertation, The University of Texas at Dallas, 2009.