Nonparametric change point estimation for survival distributions with a partially constant hazard rate

**Lifetime Data Analysis manuscript No.**
(will be inserted by the editor)

# Nonparametric change point estimation for survival distributions with a partially constant hazard rate

**Alessandra R. Brazzale · Helmut Küchenhoff · Stefanie Krügel · Tobias S. Schiergens · Heiko Trentzsch · Wolfgang Hartl**

Received: date / Accepted: date

**Abstract** We present a new method for estimating a change point in the hazard function of a survival distribution assuming a constant hazard rate after the change point and a decreasing hazard rate before the change point. Our method is based on fitting a stump regression to p-values for testing hazard rates in small time intervals. We present three real data examples describing survival patterns of severely ill patients, whose excess mortality rates are known to persist far beyond hospital discharge. For designing survival studies in these patients and for the definition of hospital performance metrics

We acknowledge support of the second author by a University of Padova Visiting Scientist Scholarship 2014.

Alessandra R Brazzale
Dipartimento di Scienze Statistiche, Università degli studi di Padova, Italy
E-mail: alessandra.brazzale@unipd.it

Helmut Küchenhoff
Statistical Consulting Unit, Department of Statistics, Ludwig-Maximilians-Universität München, Germany
E-mail: kuechenhoff@stat.uni-muenchen.de

Stefanie Krügel
Department of Statistics, Ludwig-Maximilians-Universität München, Germany
E-mail: stefanie.kruegel@gmail.com

Tobias S. Schiergens
Department of General, Visceral, Transplantation and Vascular Surgery, University School of Medicine, Grosshadern Campus, Ludwig-Maximilians-Universität München, Germany
E-mail: Tobias.Schiergens@med.uni-muenchen.de

Heiko Trentzsch
Institut für Notfallmedizin und Medizinmanagement INM, Klinikum der Universität München, Ludwig-Maximilians-Universität München, Germany
E-mail: Heiko.Trentzsch@med.uni-muenchen.de

Wolfgang Hartl
Department of General, Visceral, Transplantation, and Vascular Surgery, University School of Medicine, Grosshadern Campus, Ludwig-Maximilians-Universität München, Germany
E-mail: Wolfgang.Hartl@med.uni-muenchen.de

(e.g. mortality), it is essential to define adequate and objective end points. The reliable estimation of a change point will help researchers to identify such end points. By precisely knowing this change point, clinicians can distinguish between the acute phase with high hazard (time elapsed after admission and before the change point was reached), and the chronic phase (time elapsed after the change point) in which hazard is fairly constant. We show in an extensive simulation study that maximum likelihood estimation is not robust in this setting, and we evaluate our new estimation strategy including bootstrap confidence intervals and finite sample bias correction.

**Keywords** Change point · Survival · Hazard rate · ICU · Acute phase

## 1 Introduction

In survival studies of critically ill patients and in the reliability analysis of components, the structure of the hazard rate as a function of time is of major interest. Different phases of the hazard rate have been identified in patients who had required a therapy in an intensive care unit (ICU) (Schneider et al 2010). In the first phase after ICU admission, the hazard rate has high values and decreases rapidly over time. This period reflects the acute phase where the disease is immediately life-threatening. The second phase is defined by a constant hazard rate over time (long-term survival) and reflects a chronic phase where survivors of the acute phase have a lower risk of dying, but may still be suffering from long-term sequelae of their preceding organ failure.

The question arises, at which point in time, denoted by the change point, the second phase starts. This question is important, since it turned out in different studies that predictors for long-term and short-term survival are fundamentally different; see Schneider et al (2010) and Callcut et al (2016). Interestingly a comparable problem is encountered with specific products (e.g. automotive electronic), where different phases of hazard have been proposed when analyzing the time elapsed after sale and before warranty returns (see Kleyner and Sandborn 2005; Yang et al 2012; Altun and Comert 2016). The corresponding structures of hazard rates, and the importance of identifying change points, are comparable to those encountered in survival analyses of critically ill patients.

There is a rich literature on the estimation of change points in the hazard rate. Many papers address the problem of finding an abrupt change, i.e. a jump in the hazard function. In an early paper, Matthews and Farewell (1982) discuss the estimation of piecewise constant hazard with one change point based on non standard likelihood ratio test theory. Detailed inference for change point estimation is presented by Loader (1991). Noura and Read (1990) estimate a proportional hazard model including a change point using a Weibull model. Other approaches based on the Nelson-Aalen estimator and on least squares have been developed by Chang et al (1994) and Gijbels and Gürler (2003). Nonparametric approaches with a general hazard function with one or more non continuous change points have been addressed by Müller and

Wang (1990), Antoniadis et al (2000) and Wang et al (2014). The case of a continuous hazard rate with a change point is on the other hand rarely covered. In a recent paper, Altun and Comert (2016) use Weibull-Exponential models to represent the typical L-shaped hazard rates of electronic products. While the latter authors make use of distributional assumptions, the first nonparametric approach with a Bayesian estimation strategy using a stochastic jump process model was proposed by Yang et al (2012).

In this paper, we focus on the case of a hazard function with one change point and a constant hazard after the change point; the solution of the problem is motivated by our recent observation of changing hazard rates of ICU patients. In Section 2, some notation is introduced and maximum likelihood estimation of the change point is discussed. The new procedures for change point estimation using p-values are presented in Section 3. A numerical assessment of the different approaches by an extensive simulation study is presented in Section 4. In Section 5, the results for three studies on the survival on ICU patients are discussed followed by some discussion and an outlook in Section 6.

## 2 Definitions and settings

We are interested in survival times with a distribution function $F(t)$ having a hazard rate $h(t)$, which is a function of time $t > 0$ and has the following form

$$h(t) = h_1(t) \cdot I(t < \tau) + \lambda \cdot I(t \geq \tau). \tag{1}$$

Here, $\tau > 0$ is the change point and $I(\cdot)$ denotes the indicator function. In the first phase $(0; \tau)$, the hazard function is a continuous smooth function of $t$, denoted by $h_1(t)$, while in the second phase $[\tau, \infty)$, the hazard has the constant value $\lambda > 0$. Then, the distribution function $F(t)$ becomes

$$F(t) = F_1(t) \cdot I(t < \tau) + F_1(\tau) \cdot \{1 - \exp[-\lambda(t - \tau)]\} \cdot I(t \geq \tau) \tag{2}$$

$$\text{with} \quad F_1(t) = 1 - \exp\left(-\int_0^t h(s)ds\right).$$

The assumption of a continuous hazard function is framed by $h_1(\tau) = \lambda$. Furthermore, we allow for non informative right censoring at $c_i > 0$ while we observe $n$ independent data pairs $(y_i, d_i)$, $i = 1, \ldots, n$, where $y_i = \min(t_i, c_i)$ and $d_i$ denotes the event indicator, i.e. $d_i = 0$ for censored data $(t_i > c_i)$ and $d_i = 1$ for events $(t_i \leq c_i)$.

### 2.1 Maximum likelihood

We are mainly interested in estimating the unknown change point $\tau$. As a first approach, we assume a parametric model for the first phase of the hazard

function. A standard assumption is a Weibull type hazard with parameters $\eta > 0$ and $\gamma > 0$, i.e.

$$h(t) = \frac{\gamma}{\eta}\left(\frac{t}{\eta}\right)^{(\gamma-1)} \cdot I(t < \tau) + \lambda \cdot I(t \geq \tau);$$

see e.g. Li et al (2013).

We denote by $f_1(t)$ the density and by $F_1(t)$ the distribution function of the Weibull distribution which represents the first phase. Both functions depend on the parameters $\gamma$ and $\eta$. The log likelihood of the observed data $(y_i, d_i)$, $i = 1, \ldots, n$ is given by

$$L(\gamma, \eta, \lambda, \tau) = \sum_{i=1}^{n}(l_{1i} + l_{2i} + l_{3i} + l_{4i})$$

with the four components

$$\begin{aligned}
l_{1i} &= d_i \cdot I(y_i < \tau) \log f_1(y_i; \gamma, \eta) \\
l_{2i} &= (1 - d_i) \cdot I(y_i < \tau) \log F_1(y_i; \gamma, \eta) \\
l_{3i} &= d_i \cdot I(y_i \geq \tau) \left[\log F_1(\tau; \gamma, \eta) + \log(\lambda) - \lambda(y_i - \tau)\right] \\
l_{4i} &= (1 - d_i) \cdot I(y_i \geq \tau) \left(\log F_1(\tau; \gamma, \eta) + \log\left\{1 - \exp\left[-\lambda(y_i - \tau)\right]\right\}\right)
\end{aligned}$$

defined by the status of censoring and the phase corresponding to the observed time $y_i$.

Maximum likelihood (ML) estimation can be performed by using the profile likelihood for $\tau$ and a grid search on the same parameter. The assumption of a continuous hazard function implies the following parameter restriction

$$\frac{\gamma}{\eta}\left(\frac{\tau}{\eta}\right)^{(\gamma-1)} = \lambda;$$

see also Kleyner and Sandborn (2005). An alternative assumption is a constant hazard in the first phase, i.e. $h_1(t) = \lambda_1$; see Li et al (2013). Further parametric assumptions using different parametric distributions for modelling the survival data in the first phase can be made.

### 2.2 Robustness of the ML approach

Indeed, all parametric approaches turned out to be rather unstable in our practical experience and in our simulation studies; for details on the latter see Section 4.1.1. In particular, the estimation of the change point was highly dependent on the parametric assumptions on the first phase of the hazard function. Furthermore, the likelihood turned out to be flat and the ML estimate did not exist for our real data examples; see Section 5. Therefore, other procedures based on statistical tests have been proposed; see Schneider et al (2010) for a strategy using p-values of tests for constant hazard. A new approach without any parametric assumption about the hazard function in Phase 1 is presented in the next section.

## 3 Nonparametric change point estimation

### 3.1 Change point estimation using p-values

We assume a two phase hazard function $h(t)$, as defined in (1), with a constant hazard rate in the second phase, i.e. $h(t) = \lambda$ for $t \geq \tau$. Furthermore, we assume that the hazard rate of the first phase $h_1(t)$ is higher than $\lambda$. Typical shapes of such a hazard function are displayed in Figure 1 of Section 4.

We present a strategy for the estimation of the unknown change point $\tau$ based on p-values, which was inspired by the proposal for regression models of Mallik et al (2011). As a first step, we calculate a consistent estimate $\hat{\lambda}$ for $\lambda$. The second step consists of performing a series of statistical tests on suitable intervals $(a_{k-1}; a_k]$ with the null hypothesis $H_0 : h(t) = \hat{\lambda}$ for $t \in (a_{k-1}; a_k]$ with $k = 1, \ldots, K$, against the alternative that $h(t) > \hat{\lambda}$. Since the null hypothesis is approximately true for $a_{k-1} \geq \tau$, the corresponding p-values, $pv_k$, approximately follow a uniform distribution on $(0; 1)$. If $a_{k-1} < \tau$, the null hypothesis is not true and the corresponding p-values converge to 0 in probability for increasing sample size. In the last step, these two properties of the p-values are exploited to fit a two phase piecewise constant regression model (a so-called stump) to the p-values as outcome variable. The change point estimate of the regression model for the p-values gives the estimate of the change point $\tau$ of the hazard function.

More specifically, our estimation strategy for the change point $\tau$ given i.i.d. observations $(y_i, d_i), i = 1 \ldots, n$, is as follows.

1. We first calculate a consistent estimate $\hat{\lambda}$ for $\lambda$. We assume that an upper limit $\tau_{max}$ for the change point is known, i.e. $\tau \leq \tau_{max}$. Since the hazard function is constant for $t > \tau_{max}$, the conditional distribution of $(T_i - \tau_{max}) \mid (T_i > \tau_{max})$ is exponential. A consistent estimate for $\lambda$ is the maximum likelihood estimate for the rate parameter of an exponential distribution calculated on the observations with $y_i > \tau_{max}$, i.e.

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} I(y_i > \tau_{max}) \cdot d_i}{\sum_{i=1}^{n} I(y_i > \tau_{max}) \cdot (y_i - \tau_{max})}. \tag{3}$$

2. We define a grid $a_0 < a_1 < \cdots < a_K = \tau_{max}$ with fixed grid length $l$, i.e $a_k = a_0 + k \cdot l$ for $k = 1, \ldots, K$. The number of events in the interval $(a_{k-1}; a_k]$ is denoted by $X_k$, $k = 1, \ldots, K$. If the hazard in the $k$th interval is equal to $\lambda$, then $X_k$ follows a binomial distribution with parameters $N_k = \sum_{i=1}^{n} I(y_i > a_{k-1})$ and $pr_k = 1 - \exp\{-\lambda(a_k - a_{k-1})\}$. Hence, we perform a one-sided exact binomial test for the null hypothesis $H_{0k} : pr_k = 1 - \exp\{-\hat{\lambda}(a_k - a_{k-1})\}$. The corresponding p-value is

$$pv_k = 1 - F_{bin}(X_k; N_k, \hat{pr}_k), \tag{4}$$

where $F_{bin}$ denotes the binomial distribution function. If there are censored observations in the interval $(a_{k-1}; a_k]$, we correct the denominator of the

binomial distribution as

$$N_k^c = \left[ N_k - \sum_{i=1}^{n} I(a_{k-1} < y_i \leq a_k) \cdot I(d_i = 0) \cdot \frac{y_i - a_{k-1}}{a_k - a_{k-1}} \right],$$

where the value in square brackets is rounded to the closest integer number.

3. Was $\lambda$ exactly equal to $\hat{\lambda}$, the p-values in (4) calculated for $a_{k-1} \geq \tau$ would follow a uniform distribution on $(0;1)$ with expected value larger than 0.5 because of the discreteness of the test statistic. For the intervals with $a_{k-1} < \tau$, the binomial probability is higher than $\hat{pr}_k$, i.e. the corresponding p-values should be close to 0. To account for both, the estimation of $\lambda$ and the discreteness of the test statistic, we use a two phase piecewise constant regression function, $E(pv_k) = \beta \cdot I(a_{k-1} > \tau)$, with an unknown constant $\beta > 0$ to model the p-values. The estimate of the change point $\tau$ for the fixed grid $a_0 < a_1 < \cdots < a_K = \tau_{max}$ is given by

$$\hat{\tau}_e = \arg\min_{\tau} \left\{ \min_{\beta} \sum_{k=1}^{K} \left[ pv_k - \beta \cdot I(a_{k-1} > \tau) \right]^2 \right\}. \tag{5}$$

   An further alternative, we evaluated, is to replace the p-values calculated at (4) with their mid-p values, whose expected values are exactly 0.5 if $\lambda = \hat{\lambda}$. The corresponding stump regression model, $E(pv_k) = 0.5 \cdot I(a_{k-1} > \tau)$, is the one discussed in Mallik et al (2011) though for a rather different setting. As we will show in our simulation study (see Section 4), this alternative does not improve over our original proposal.

4. Expression (5) is based on a grid with constant grid length $l$. Since the result depends on the starting point $a_0$, we repeat the estimation process for the starting points $a_0, a_0 + \delta/l, \ldots, a_0 + (J-1) \cdot (\delta/l)$, with $\delta = l/J$ for a suitable integer $J$. We use as estimate for $\tau$ the value which gives the best fit, i.e.

$$\hat{\tau} = \arg\min_{\tau} \left( \min_{j \in 0,\ldots,J-1} \left\{ \min_{\beta} \sum_{k_j=1}^{K} \left[ pv_{k_j} - \beta \cdot I(a_{k-1} + j \cdot \delta > \tau) \right]^2 \right\} \right), \tag{6}$$

   where $pv_{k_j}$ identifies the p-value obtained for the $k$th interval of the grid which starts at $a_0 + j \cdot (\delta/l)$, for $j = 0, \ldots, J-1$. So, the grid length $l$ is a tuning parameter to be chosen.

The performance of our nonparametric change point estimator is assessed numerically in Section 4.2. A fitting routine is provided in the R package `CPsurv` which is freely available on CRAN.

### 3.2 Bootstrap inference

To account for possible finite sample bias in the change point estimator defined by (6) for $\tau$, we propose a bias correction based on a nonparametric bootstrap

resampling; see Efron and Tibshirani (1993, Chapter 10). We generate $B$ bootstrap samples $\{y_b^*, d_b^*\}$ of size $n$ from the estimated distribution $\hat{F}(t)$, where $\{y_b^*, d_b^*\} = \{(y_{b1}^*, d_{b1}^*), \ldots, (y_{bn}^*, d_{bn}^*)\}$, for $b = 1, \ldots, B$. The change point $\tau$ in (2) is replaced by (6) and the constant hazard $\lambda$ in the second phase by $\hat{\lambda}$. The distribution function of the first phase, where $t < \tau$, is estimated by calculating the Kaplan-Meier estimator for our data; this yields a consistent estimator of $F_1(t)$. To mimic the noninformative censoring mechanism, we censor our bootstrap data using an independent exponential distribution. The corresponding rate parameter is chosen such that the expected number of censored observations in the bootstrap samples is in line with the observed number of censored observations of our sample. We then apply (6) to the bootstrap samples which results in $B$ bootstrap estimates $\hat{\tau}_b^*$, $b = 1, \ldots, B$. The bias corrected estimate of the change point $\tau$ is given by

$$\hat{\tau}_{bc} = 2\hat{\tau} - \text{median}(\hat{\tau}_b^*)_{b=1}^{B}; \tag{7}$$

see Section 3.9.1 of Davison and Hinkley (1997). We call this estimator nonparametric, since the first part of the distribution function $F(t)$ is estimated by the nonparametric Kaplan-Meier estimator.

Confidence intervals for $\tau$ are also obtained by using the previously described nonparametric bootstrap procedure for censored data. We checked the suitability of both, the symmetric normal approximation confidence interval and the percentile interval (Davison and Hinkley 1997, Chapter 5). A nested bootstrap can further be used to include median bias correction. This nested bootstrap procedure is computer intensive but still feasible.

The performance of the bias corrected nonparametric estimate of $\tau$ will be numerically assessed in Section 4.3, while the real coverage of the bootstrap confidence intervals is investigated in Section 4.2. Bias corrected estimation and the computation of confidence intervals is provided by the R package `CPsurv`.

### 3.3 Tuning parameter

As mentioned at the end of Section 3.1, the grid length $l$ is a tuning parameter which governs the performance of our nonparametric estimator $\hat{\tau}$ for the change point $\tau$. The simulation study reported in Section 4.4 evaluates different choices for the grid length $l$. In fact, the grid length must be chosen in such a way, that the expected number of events within an interval is not too low. This is required to provide enough power to detect a possible small change, even though the chosen binomial test is the uniformly most powerful test for the given scenario provided that $\lambda = \hat{\lambda}$. Fixing a too small interval width could lead to a low number of events within the intervals preceding the true change point $\tau$. And this in turn could cause a downwards bias in the change point estimate because of the false negatives. On the other hand, a wide interval

possibly masks the change point. In Section 5, we present a number of graphical tools for the choice of the tuning parameter, while providing an extensive discussion of this aspect.

The grid length $l$ furthermore plays a central role in determining the consistency of our nonparametric estimator $\hat{\tau}$. Having defined $N_1, \ldots, N_K$ as the number of observations at risk in the $K$ intervals $(a_{k-1}; a_k]$, for $i = 1, \ldots, K$, (see Step 2 of the algorithm of Section 3.1), let $N_{K+1}$ further represents the same quantity for the interval $(a_K, \infty)$. Let's further suppose that the change point $\tau$ falls on one extreme of the intervals, say $a_{k_0}$. If $\min(N_1, \ldots, N_{K+1})$ tends to infinity, the consistent estimator $\hat{\lambda}$ converges in probability to $\lambda$ and $\hat{pr}_k \to pr_k$. Then, the $p$-values $pr_k$ converge to 0 for $k < k_0$ and the expected value of the p-values on the right of $\tau$ coverges to $1/2$. The least square estimator of $\beta$ converges to the true expected value of the p-values on the right of $\tau$, and this in turn implies $\hat{\tau}_e$ to converge to the true change point $\tau$. To relax the assumption that $\tau$ falls on one of the extremes defining the grid, we further assume that the grid width $(a_k - a_{k-1})$, for $k = 1, \ldots, K$, shrinks to zero, while $K \to \infty$, so that eventually $\tau$ will fall on one of the extremes. Consistency is guaranteed provided that the binomial probabilities $pr_k$ tend to zero at a lower rate than the interval width $l$.

## 4 Numerical assessment

We conducted an extensive simulation study for evaluating our new estimator in realistic set-ups. The first part of the study evaluates both, maximum likelihood estimation and the alternative settings of the estimating procedure presented in Section 3.1. In the second part of the simulation study, we examine our estimation strategy under different data scenarios with the focus on its general performance. In the third part, we consider strategies for bias correction. Different possible choices of the tuning parameter are evaluated in the last part of the simulation study.

Inspired by our motivating example, we generated the data from a Weibull distribution with parameters $\eta$ and $\gamma$ in the first phase and a constant hazard in the second phase, where $t \geq \tau$. We considered two main cases, a continuous change in the hazard function and a relevant jump in the same at the change point. More specifically, observations are simulated according to the survival distribution characterized by the hazard function

$$h(t) = h_1(t) \cdot I(t < \tau) + \lambda \cdot I(t \geq \tau), \quad \text{where} \quad h_1(t) = \frac{\gamma}{\eta} \cdot \left(\frac{t}{\eta}\right)^{\gamma - 1},$$

with $\eta = 100$ and $\gamma = 0.44$. For the change point we evaluated the values $\tau \in \{50, 55, 90, 100\}$. In the continuous setting, the constant hazard after the change point was given by $\lambda = h_1(\tau)$. In the second scenario we set the hazard after the change point to $\lambda = h_1(\tau)/2$, which causes a drop in the hazard rate at $\tau$. Two template hazard functions for both scenarios are given in Figure 1 for $\tau = 50$ and $\tau = 90$. As upper bound for $\tau$ we used, depending on the setting,

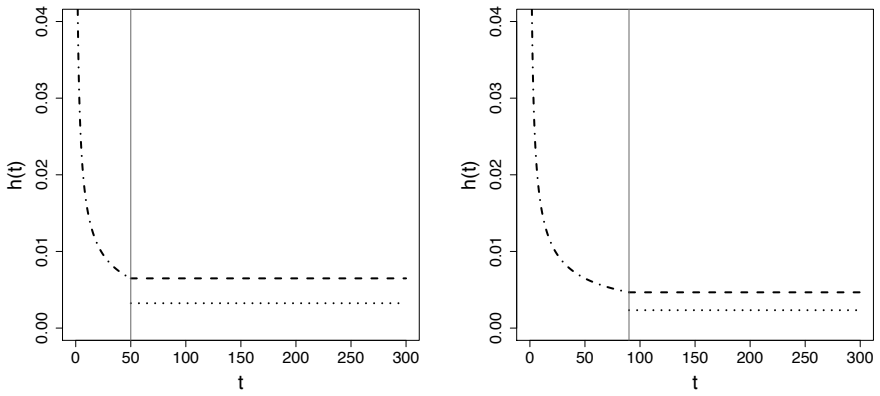


**Fig. 1** Two template hazard functions of the assessed simulation scenarios for the change point values $\tau = 50$ (on the left) and $\tau = 90$ (on the right). The dotted line represents the Weibull-Exponential two phase model for which the hazard function drops at the change point, while the dashed line corresponds to the same model but with a continuous change in the hazard at $\tau$. Both models are characterized by the same $h_1(t)$ hazard function in the first phase, i.e. where $t < \tau$.

$\tau_{max} = 200$ if $\tau \in \{50, 55\}$ and $\tau_{max} = 360$ if $\tau \in \{90, 100\}$. Two different sample sizes were assessed, $n = 1,000$ and $n = 5,000$. To summarize the results of the simulation study we considered as summary measures the mean, the median, the mean average distance (MAD) and the root mean squared error (RMSE) of 1,000 simulations.

## 4.1 Assessed estimation strategies

In the first run of the simulation study, we checked which of the following estimation strategies performs best:

1. ML estimation as presented in Section 2.1 for exact and rounded data using a grid search.
2. our new estimator $\hat{\tau}$ under alternative settings. The latter are defined by combining the following options (see Section 3.1):
   - the two types of regression model for the p-values, i.e. $E(pv_k) = \beta \cdot I(a_{k-1} > \tau)$ and $E(pv_k) = 0.5 \cdot I(a_{k-1} > \tau)$;
   - the use of common p-values or of mid-p values;
   - different values for the grid length $l$.

### *4.1.1 Maximum likelihood estimation*

The results presented in the upper half of Table 1 show that maximum likelihood (ML) estimation gives the best results when we use the exact data. This

**Table 1** Simulation results for the Weibull-Exponential two phase model and change point $\tau$. Randomly censored survival times with a continuous hazard rate were generated. The results for maximum likelihood (ML) estimation for exact and rounded data are shown in the upper table. The results in the lower table consider several settings for our new estimator $\hat{\tau}$ for exact data. Here, $n$ denotes the sample size and $l$ the grid length. Flexible $\beta$ relates to a flexible stump regression model, while $\beta = 0.5$ indicates a stump regression model with $\beta = 0.5$. The median, mean, mean average distance (MAD) and root mean squared error (RMSE) for 1,000 repetitions are presented.

| setting | | | ML estimation | | | |
|---|---|---|---|---|---|---|
| $n$ | $\tau$ | data | median | mean | MAD | RMSE |
| 1000 | 50 | exact | 46 | 47 | 7.57 | 9.31 |
| 1000 | 55 | exact | 51 | 52 | 7.50 | 9.78 |
| 1000 | 90 | exact | 86 | 88 | 38.42 | 42.71 |
| 1000 | 100 | exact | 96 | 99 | 49.22 | 53.92 |
| 1000 | 50 | rounded | 87 | 91 | 41.26 | 47.65 |
| 1000 | 55 | rounded | 95 | 99 | 49.41 | 55.56 |
| 1000 | 90 | rounded | 147 | 159 | 109.2 | 119.03 |
| 1000 | 100 | rounded | 162 | 176 | 125.57 | 135.96 |

| setting | | nonparametric estimation | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $n$ | $\tau$ | test | regression | $l$ | median | mean | MAD | RMSE |
| 1000 | 50 | p-value | $\beta = 0.5$ | 10 | 40 | 48 | 19.04 | 29.46 |
| 1000 | 50 | mid-p | $\beta = 0.5$ | 10 | 40 | 52 | 20.99 | 33.42 |
| 1000 | 50 | p-value | flexible $\beta$ | 10 | 41 | 46 | 15.97 | 22.06 |
| 1000 | 50 | mid-p | flexible $\beta$ | 10 | 41 | 46 | 16.09 | 22.78 |
| 1000 | 50 | p-value | $\beta = 0.5$ | 20 | 42 | 63 | 27.77 | 45.26 |
| 1000 | 50 | mid-p | $\beta = 0.5$ | 20 | 45 | 67 | 31.38 | 51.07 |
| 1000 | 50 | p-value | flexible $\beta$ | 20 | 46 | 60 | 24.18 | 38.99 |
| 1000 | 50 | mid-p | flexible $\beta$ | 20 | 45 | 60 | 25.11 | 40.99 |

finding is in line with the known large sample properties of general likelihood theory, despite the fact that we are not in a regular setting as the likelihood function is not differentiable in $\tau$. However, rounding on full days, which represents only a slight, though rather common, modification of the data in medical frameworks, causes a breakdown in the performance of the ML estimator. This is supported by our experience with practical data, where we faced convergence problems and instability issues using the likelihood function. Additional simulations using ML estimation with a misspecified parametric distribution in the first phase also yielded high values of MAD and RMSE (results not shown).

#### *4.1.2 Settings for the new estimator*

We present only results using the exact data, since there was basically no difference with rounded data for the new estimator. The results in the second part of Table 1 indicate that using a flexible value for the $\beta$ coefficient reduces the MAD and the RMSE compared to when $\beta = 0.5$. Furthermore, using the mid-p value does not improve the estimator based on standard p-values. Therefore, the following runs of the simulation study will focus on the setting where $\beta$ is estimated using standard p-values. The choice of the grid length $l$,

which is the main tuning parameter in our procedure, has a rather high impact on the performance of the estimator. In our simulation scenarios, a small value ($l$=10) yielded much better results than the larger width of 20. We will further explore this aspect in the last part of the simulation study (Section 4.4).

## 4.2 Main evaluation

For our main evaluation, we assumed different change points, and different censoring scenarios. Under all scenarios, we evaluated our nonparametric change point estimator using a 10 days grid length. As will be shown in Section 4.4, the best results are achieved with this choice of the tuning parameter. Our scenarios considered the change point values $\tau \in \{50, 55, 90, 100\}$, and three types of right censoring (random censoring, Type I censoring at $t = 540$ and no censoring). The amount of censored observations in the data ranged between 5% and 20%; we won't further comment these rates as they didn't influence the performance of our nonparametric estimator. The sample size was $n \in \{1,000, 5,000\}$. The results are presented in Table 2. Our estimator performs much better, i.e. exhibits a much smaller MAD, for the scenario with a jump. This is not surprising, since the information provided by the data for the change point is obviously higher. Furthermore, the MAD for the scenarios with a change point at $\tau \in \{90, 100\}$ is higher than for those with a change point at $\tau \in \{50, 55\}$. This is due to the structure of the hazard rate; see Figure 1, where the change point is more pronounced at $\tau \in \{50, 55\}$. If we focus on the practically more relevant case of no jump in the hazard function, there is a substantial reduction of the MAD when increasing the sample size. The coverage rate for both types of bootstrap confidence intervals is acceptable for the higher sample size. The coverage rate is low for the difficult case when $\tau \in \{90, 100\}$ and $n = 1,000$. For the continuous hazard case, there is no substantial bias for $\tau \in \{50, 55\}$, but for $\tau \in \{90, 100\}$ there seems to be a bias even for the high sample size of 5,000. Therefore, a bias correction procedure was developed and assessed.

## 4.3 Bias correction

We checked our bias correction procedure from Section 3.2 for the scenario with no jump. Since the results were similar for the different censoring schemes we present only results for the case of no censoring; see Table 3. The bias correction reduces the bias, but MAD and RMSE are increased. This bias-variance trade-off was observed under all other scenarios (results not presented). Therefore, a bias correction can be useful, but it should not be applied as the standard procedure.

**Table 2** Simulation results for the Weibull-Exponential two phase model, with and without a jump in the hazard rate at the change point $\tau$. The data were generated using different censoring types. The simulation was run with the R package `CPsurv (v1.0.0)`. Two sample sizes were considered, $n = 1,000$ and $n = 5,000$, for the grid length $l = 10$. The median, mean, mean average distance (MAD) and root mean squared error (RMSE) for 1,000 repetitions are presented. The average (ø) length and real coverage (cover) are given for both, the normal approximation and the percentile confidence intervals.

| setting | | | estimation | | | | ci normal | | ci percentile | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $\tau$ | $l$ | median | mean | MAD | RMSE | ø length | cover | ø length | cover |
| *no censoring; continuous hazard* | | | | | | | | | | |
| 1000 | 50 | 10 | 41 | 44 | 13.63 | 17.82 | 66 | 0.90 | 66 | 0.92 |
| 1000 | 55 | 10 | 44 | 47 | 14.54 | 18.36 | 66 | 0.88 | 66 | 0.90 |
| 1000 | 90 | 10 | 68 | 72 | 24.31 | 28.30 | 84 | 0.75 | 81 | 0.77 |
| 1000 | 100 | 10 | 74 | 77 | 27.39 | 31.34 | 85 | 0.69 | 81 | 0.73 |
| 5000 | 50 | 10 | 46 | 51 | 10.41 | 16.54 | 67 | 0.96 | 65 | 0.98 |
| 5000 | 55 | 10 | 52 | 55 | 10.23 | 15.85 | 66 | 0.96 | 63 | 0.97 |
| 5000 | 90 | 10 | 80 | 83 | 14.90 | 19.17 | 80 | 0.91 | 76 | 0.93 |
| 5000 | 100 | 10 | 87 | 90 | 16.18 | 19.88 | 80 | 0.88 | 75 | 0.89 |
| *no censoring; hazard with jump at $\tau$* | | | | | | | | | | |
| 1000 | 50 | 10 | 54 | 57 | 8.36 | 14.17 | 52 | 0.94 | 49 | 0.96 |
| 1000 | 55 | 10 | 60 | 62 | 8.27 | 13.91 | 52 | 0.95 | 49 | 0.98 |
| 1000 | 90 | 10 | 93 | 96 | 8.32 | 14.61 | 54 | 0.95 | 51 | 0.98 |
| 1000 | 100 | 10 | 103 | 106 | 7.98 | 13.27 | 54 | 0.96 | 51 | 0.99 |
| 5000 | 50 | 10 | 56 | 59 | 9.61 | 15.39 | 53 | 0.91 | 48 | 0.89 |
| 5000 | 55 | 10 | 61 | 65 | 9.78 | 15.21 | 52 | 0.92 | 48 | 0.87 |
| 5000 | 90 | 10 | 95 | 99 | 8.85 | 14.51 | 54 | 0.93 | 48 | 0.93 |
| 5000 | 100 | 10 | 105 | 109 | 8.83 | 14.78 | 53 | 0.93 | 48 | 0.93 |
| *random censoring; continuous hazard* | | | | | | | | | | |
| 1000 | 50 | 10 | 41 | 46 | 15.97 | 22.06 | 77 | 0.90 | 78 | 0.93 |
| 1000 | 55 | 10 | 44 | 49 | 16.61 | 22.49 | 77 | 0.88 | 77 | 0.91 |
| 1000 | 90 | 10 | 67 | 74 | 28.86 | 35.99 | 119 | 0.82 | 119 | 0.84 |
| 1000 | 100 | 10 | 72 | 79 | 31.73 | 37.69 | 118 | 0.78 | 116 | 0.81 |
| 5000 | 50 | 10 | 46 | 52 | 12.28 | 20.06 | 78 | 0.96 | 77 | 0.97 |
| 5000 | 55 | 10 | 51 | 56 | 11.93 | 19.75 | 76 | 0.95 | 74 | 0.96 |
| 5000 | 90 | 10 | 78 | 85 | 20.10 | 28.74 | 112 | 0.90 | 108 | 0.92 |
| 5000 | 100 | 10 | 86 | 92 | 20.88 | 28.45 | 109 | 0.89 | 105 | 0.90 |
| *random censoring; hazard with jump at $\tau$* | | | | | | | | | | |
| 1000 | 50 | 10 | 54 | 58 | 9.19 | 17.49 | 59 | 0.94 | 56 | 0.96 |
| 1000 | 55 | 10 | 60 | 63 | 9.24 | 17.18 | 58 | 0.96 | 55 | 0.98 |
| 1000 | 90 | 10 | 93 | 98 | 10.48 | 19.60 | 72 | 0.96 | 67 | 0.98 |
| 1000 | 100 | 10 | 103 | 106 | 9.42 | 16.48 | 70 | 0.97 | 66 | 0.99 |
| 5000 | 50 | 10 | 55 | 60 | 9.83 | 16.97 | 59 | 0.94 | 54 | 0.90 |
| 5000 | 55 | 10 | 61 | 65 | 9.88 | 15.39 | 58 | 0.94 | 53 | 0.89 |
| 5000 | 90 | 10 | 95 | 99 | 9.68 | 17.09 | 69 | 0.96 | 62 | 0.93 |
| 5000 | 100 | 10 | 105 | 110 | 10.76 | 18.71 | 67 | 0.95 | 60 | 0.94 |
| *type I censoring; continuous hazard* | | | | | | | | | | |
| 1000 | 50 | 10 | 41 | 45 | 13.74 | 18.26 | 68 | 0.91 | 69 | 0.93 |
| 1000 | 55 | 10 | 44 | 48 | 14.87 | 19.68 | 69 | 0.88 | 68 | 0.9 |
| 1000 | 90 | 10 | 68 | 70 | 23.77 | 27.17 | 74 | 0.72 | 71 | 0.76 |
| 1000 | 100 | 10 | 73 | 76 | 27.19 | 30.90 | 76 | 0.66 | 73 | 0.72 |
| 5000 | 50 | 10 | 46 | 51 | 10.45 | 16.33 | 69 | 0.96 | 67 | 0.98 |
| 5000 | 55 | 10 | 52 | 56 | 10.73 | 17.04 | 68 | 0.96 | 66 | 0.97 |
| 5000 | 90 | 10 | 79 | 82 | 14.83 | 18.34 | 67 | 0.87 | 64 | 0.91 |
| 5000 | 100 | 10 | 87 | 90 | 16.66 | 20.20 | 67 | 0.84 | 64 | 0.89 |
| *type I censoring; hazard with jump at $\tau$* | | | | | | | | | | |
| 1000 | 50 | 10 | 54 | 58 | 8.55 | 15.08 | 55 | 0.95 | 52 | 0.96 |
| 1000 | 55 | 10 | 60 | 62 | 8.45 | 14.66 | 55 | 0.95 | 51 | 0.98 |
| 1000 | 90 | 10 | 93 | 96 | 8.18 | 14.37 | 52 | 0.95 | 50 | 0.98 |
| 1000 | 100 | 10 | 103 | 106 | 7.85 | 13.22 | 52 | 0.96 | 50 | 0.99 |
| 5000 | 50 | 10 | 56 | 59 | 9.35 | 15.05 | 56 | 0.93 | 51 | 0.89 |
| 5000 | 55 | 10 | 61 | 65 | 9.94 | 15.94 | 55 | 0.92 | 50 | 0.87 |
| 5000 | 90 | 10 | 95 | 99 | 8.90 | 15.32 | 50 | 0.92 | 46 | 0.91 |
| 5000 | 100 | 10 | 105 | 109 | 8.96 | 15.02 | 49 | 0.92 | 45 | 0.94 |

**Table 3** Simulation results for the Weibull-Exponential two phase model with a continuous hazard at the change point $\tau \in \{55, 90\}$. The generated data are neither rounded nor censored. The simulation was run with the R package `CPsurv (v1.0.0)`. Two sample sizes were considered, $n = 1,000$ and $n = 5,000$, for the grid length $l = 10$. The median, mean, mean average distance (MAD) and root mean squared error (RMSE) for 1,000 repetitions are presented.

| setting | | | raw estimate | | | | bias corrected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $\tau$ | $l$ | median | mean | MAD | RMSE | median | mean | MAD | RMSE |
| 1000 | 55 | 10 | 44 | 47 | 14.54 | 18.36 | 44 | 50 | 17.21 | 24.42 |
| 1000 | 90 | 10 | 68 | 72 | 24.31 | 28.30 | 70 | 77 | 27.16 | 35.21 |
| 5000 | 55 | 10 | 52 | 55 | 10.23 | 15.85 | 52 | 57 | 13.47 | 23.01 |
| 5000 | 90 | 10 | 80 | 83 | 14.90 | 19.17 | 80 | 87 | 18.19 | 27.44 |

**Table 4** Simulation results for the Weibull-Exponential two phase model with a continuous hazard at the change point $\tau \in \{55, 90\}$. The generated data are randomly censored. The simulation was run with the R package `CPsurv (v1.0.0)`. Two sample sizes were considered, $n = 1,000$ and $n = 5,000$, for the grid length $l \in \{3, 5, 10, 20\}$. The median, mean, mean average distance (MAD) and root mean squared error (RMSE) for 1,000 repetitions are presented. The average (ø) length and real coverage (cover) are given for both, the normal approximation and the percentile confidence intervals.

| setting | | | estimation | | | | ci normal | | ci percentile | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $\tau$ | l | median | mean | MAD | RMSE | ø length | cover | ø length | cover |
| 1000 | 55 | 3 | 35 | 36 | 19.81 | 21.52 | 35 | 0.38 | 33 | 0.35 |
| 1000 | 55 | 5 | 39 | 40 | 16.78 | 19.03 | 47 | 0.64 | 45 | 0.69 |
| 1000 | 55 | 10 | 44 | 49 | 16.61 | 22.49 | 77 | 0.88 | 77 | 0.91 |
| 1000 | 55 | 20 | 49 | 62 | 24.09 | 37.56 | 109 | 0.92 | 118 | 0.97 |
| 1000 | 90 | 3 | 51 | 53 | 37.55 | 40.17 | 60 | 0.33 | 57 | 0.33 |
| 1000 | 90 | 5 | 58 | 61 | 31.78 | 35.11 | 78 | 0.56 | 74 | 0.57 |
| 1000 | 90 | 10 | 67 | 74 | 28.86 | 35.99 | 119 | 0.82 | 119 | 0.84 |
| 1000 | 90 | 20 | 75 | 92 | 37.55 | 56.02 | 177 | 0.91 | 191 | 0.93 |
| 5000 | 55 | 3 | 44 | 45 | 11.04 | 12.57 | 30 | 0.62 | 28 | 0.64 |
| 5000 | 55 | 5 | 47 | 49 | 9.55 | 11.43 | 42 | 0.85 | 40 | 0.87 |
| 5000 | 55 | 10 | 51 | 56 | 11.93 | 19.75 | 76 | 0.95 | 74 | 0.96 |
| 5000 | 55 | 20 | 55 | 69 | 22.53 | 38.90 | 111 | 0.93 | 114 | 0.98 |
| 5000 | 90 | 3 | 67 | 68 | 22.44 | 24.56 | 47 | 0.48 | 44 | 0.44 |
| 5000 | 90 | 5 | 72 | 74 | 18.98 | 22.11 | 64 | 0.73 | 61 | 0.73 |
| 5000 | 90 | 10 | 78 | 85 | 20.10 | 28.74 | 112 | 0.90 | 108 | 0.92 |
| 5000 | 90 | 20 | 84 | 102 | 30.22 | 51.87 | 177 | 0.94 | 183 | 0.97 |

### 4.4 Tuning parameter

We focused on the case of a continuous hazard function and checked different choices for the grid length $l$. The estimators were compared for $l \in \{3, 5, 10, 20\}$, each for $\tau \in \{55, 90\}$ and $n \in \{1,000, 5,000\}$ using randomly censored data. The results are presented in Table 4.4. The RMSE and MAD were similar for the 10 days grid, while there was a substantial loss in the performance of the estimator for $l = 20$. Furthermore, the real coverage of confidence intervals was notably below the nominal value for the small interval widths $l = 3, 5$. So the choice of $l = 10$ is optimal in our setting. Interestingly, this holds for both, small ($n = 1,000$) and large ($n = 5,000$) sample sizes. We obtained similar results for the other two censoring schemes.

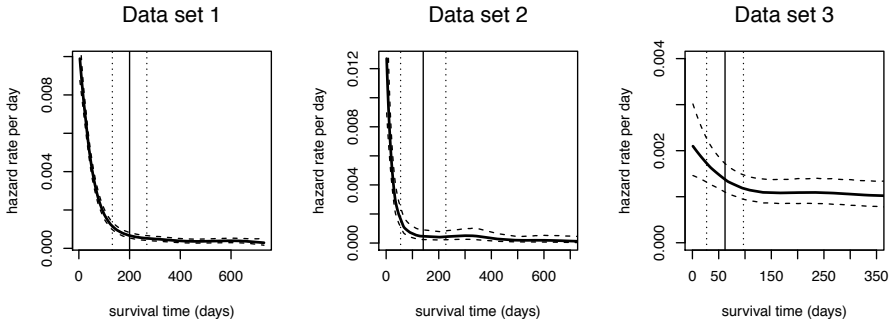


**Fig. 2** Smoothed risk function per day (bold line) with 95% confidence bands (dashed line) for patient survival. The solid and dotted vertical lines, respectively identify our nonparametric estimate of the change point and the corresponding 95% bootstrap confidence interval based on the normal approximation, obtained using a 20 days grid length. Data set 1: Survival of critically ill patients after successful acute phase therapy. Data set 2: Survival of patients after severe trauma. Data set 3: Postoperative survival of patients after a partial hepatectomy.

## 5 Application

We assessed the potential of our nonparametric change point estimator by applying it to both, the motivating example (survival times of critically ill patients) introduced in Section 1, and to two further sets of data from the authors' collaborative work.

Figure 2 displays the smoothed risk function per day for the three data sets considered: our motivating example is shown on the left, while the second and third data sets on patient survival are shown in the middle and rightmost panels. The three risk functions exhibit the typical L-shaped form of a two phase model, where the acute phase (first part) is characterized by a more or less rapid decrease of the hazard rate, which then stabilizes in the post acute phase (second part). In all three cases, the model defined in Section 2 well describes the dynamics of the survival process. However, the maximum likelihood estimate of the change point of interest does not exist in all three cases. The profile likelihood function for $\tau$, in fact, keeps increasing as the latter moves towards infinity. Furthermore, the three data sets are characterized by a decreasing information content, causing a broadening of the confidence bands for the hazard function.

All results reported in this section were obtained with the code available through our `R` package `CPsurv (v1.0.0)`. The nonparametric raw ($\hat{\tau}$) and median bias corrected ($\hat{\tau}_{bc}$) estimates are based on traditional p-values, with an unknown $\beta$ in the stump regression model. Two grid lengths were explored: 10 days and 20 days. The 95% bootstrap confidence intervals are based on 999 replicates and use the normal approximation; 49 replicates were used for median bias correction.

**Table 5** Raw ($\tau$) and median bias corrected ($\hat{\tau}_{bc}$) nonparametric change point estimates, with corresponding 95% bootstrap confidence intervals based on the normal approximation for the three examples considered. Data set 1: Survival of critically ill patients. Data set 2: Survival after severe trauma. Data set 3: Postoperative survival after partial hepatectomy. Two grid lengths were used: 10 days and 20 days. The estimate of $\beta$ in the stump regression model is also provided.

| | $l = 10$ | | | $l = 20$ | | |
|---|---|---|---|---|---|---|
| estimate | Data set 1 | Data set 2 | Data set 3 | Data set 1 | Data set 2 | Data set 3 |
| $\hat{\tau}$ | 200 | 93 | 32 | 200 | 141 | 62 |
| ci normal | (140; 260) | (37; 147) | (1; 63) | (132; 268) | (55; 227) | (27; 97) |
| $\hat{\tau}_{bc}$ | 200 | 114 | 30 | 198 | 159 | 75 |
| ci normal | (120; 280) | (28; 200) | (0; 74) | (122; 274) | (14; 304) | (28; 122) |
| $\hat{\beta}$ | 0.40 | 0.75 | 0.51 | 0.26 | 0.65 | 0.49 |

### 5.1 Data set 1: Survival of critically ill patients

The aim of this first case study was to identify the onset of the chronic phase after a successful ICU-therapy. The study was conducted in the surgical intensive care unit (ICU) of the Department of General, Visceral, Transplantation, Vascular and Thoracic Surgery, University School of Medicine, Grosshadern Campus, LMU Munich, Germany. The study period extended from March 1, 1993 till February 28, 2005, and included a total of 1,638 patients who met the primary inclusion criteria. Right censoring occurred after 730 days; 913 events (deaths) were registered. This data set was first analyzed by Schneider et al (2010), who set the starting point of the chronic phase at day 198 after ICU submission using a procedure based on the likelihood ratio test for a constant hazard rate.

Table 5 reports the raw $\hat{\tau}$ and the median bias corrected ($\hat{\tau}_{bc}$) estimates of the change point for the two grid lengths considered, together with their 95% confidence intervals. In all, 778 events were used for estimation, having set $\tau_{max} = 300$. The raw estimate and its lower and upper confidence bounds for the 20 days grid are further shown in Figure 2 as solid and dotted vertical lines. The results are in close agreement with the estimate reported in Schneider et al (2010), and they are very stable. The median bias corrected confidence bands are slightly larger.

Table 5 also reports the least squares estimates of $\beta$ for the stump regression model with $l = 10$ and $l = 20$: both are lower than the theoretically expected value of 0.5, which holds for continuous test statistics. The topmost panels in Figure 3 illustrate the p-values used to estimate the stump regression model with, superimposed, the least squares fit used for the two grid lengths. The bottom line bar plots report the percentage of events per unit at risk present in each interval.

### 5.2 Data set 2: Survival after severe trauma

The survival times of this second data set refer to patients, who had experienced a severe trauma and had been admitted between 2005 and 2010 to a level I trauma center (Hospital of the University of Munich, Grosshadern and Innenstadt Campus, LMU Munich, Germany). Inclusion criteria were primary admission via trauma room following traumatic injury with a consecutive ICU stay of two or more days. A total of 543 patients met the inclusion criteria; 89 deaths were recorded. The survival time ranged from 2 to 3,401 days. Right censoring is assumed to have occurred randomly. Most studies on mortality following a severe trauma have been restricted to in-hospital stay or 30-day mortality. There is, however, growing evidence that the risk of dying after a severe trauma injury will not drop to the baseline value of the general population even for those who have survived up to hospital discharge (Eriksson et al 2016). The aim of this second study was to estimate the point in time, where the acute phase (phase of excess mortality) ends.

We set $\tau_{max} = 700$ days, which roughly corresponds to a two years period. In all, 77 events out of the 89 present in the data set contributed to the estimation of $\tau$. The raw ($\hat{\tau}$) and median bias corrected ($\hat{\tau}_{bc}$) estimates for the two considered grid lengths are given in Table 5, together with the least squares fit of $\beta$. The raw estimate for the 10 days grid amounts to day 93 (3 months), which identifies the end of the acute phase with a margin of error of $\pm$ 54 days, while the median bias corrected estimate raises this value to somewhat less than 4 months, i.e. $114 \pm 86$ days. In both cases, the commonly used reference value for the end point (day 30 after the injury) lies only marginally within the 95% confidence interval supporting the hypothesis that short follow-up times (e.g. 28-day mortality) only poorly reflect the true survival of patients after a severe trauma. When using a 20 days grid, these estimates are further enlarged by around one and a half months, setting $\tau$ to, respectively, $141 \pm 86$ days and $159 \pm 145$ days. There is, however, a larger uncertainty as shown by the wider confidence intervals. Graphical inspection of the smoothed risk function per day for this data set (middle panel of Figure 2) and of the percentages of events for unit at risk (not shown here) supports the raw estimate obtained from the 20 days grid length.

### 5.3 Data set 3: Postoperative survival after partial hepatectomy

The issue in this analysis is to correctly identify the end of the immediate postoperative period (IPP) after a partial hepatectomy. For the first analysis of this data set (Schiergens et al 2015), we had used a preliminary version of the estimation procedure presented in this paper, which stopped at Step 3 (see p. 6) and didn't account for possible finite sample bias. Confidence intervals were obtained by nonparametric bootstrap resampling. In particular, we had been able to show that after liver resection for primary and secondary malignancies, 90-day rather than the commonly used 30-day or in-hospital mortality

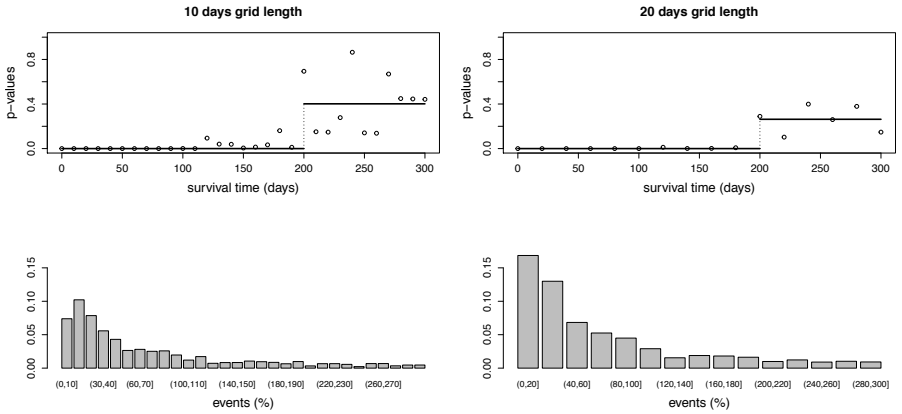


**Fig. 3** p-values used to estimate the stump regression model with superimposed least squares fit (top) and percentage of events per unit at risk for each interval (bottom) for Data set 1 on survival of critically ill patients after a successful acute phase therapy. The estimation used 778 events out of the 913 present, for a total of 1,638 observations, and $\tau_{max} = 300$. Two grid lengths were explored: 10 days and 20 days.

should be used to avoid underreporting of deaths. The original analysis had been based on 784 of 1,032 originally analyzed patients, who all had undergone elective liver resection with a curative intent between 2003 and 2013. A total of 95 patients with benign tumors had been excluded and further 153 patients could not be included in the analysis because of missing data. The change point had been estimated to be at day 80 after surgery, with a 95% confidence interval of $(40; 100)$ days. The rightmost panel of Figure 2 reports the smoothed, unadjusted daily hazard rate of the study population up to one year.

The results of our re-analysis are given in Table 5. We set $\tau_{max} = 200$ and assumed censoring to have occurred randomly; 93 out of the original 95 events were retained for estimation. The change point estimates based on a 20 days grid are in line with the findings of Schiergens et al (2015): the raw estimate amounts to 62 days with 95% confidence interval $(27; 92)$ days, while the corresponding values, when median bias correction is adopted, are 75 and $(28; 122)$ days. Both confidence intervals only marginally include the commonly used 30 days reference value. However, the results are different, when a 10 days grid is used. Long-term survival was estimated to start at day 32 $(1; 63)$ or at day 30 $(0; 74)$. The latter result would be in agreement with the commonly used 30 days reference end point of the acute postoperative phase, though it may be due to the high amount of uncertainty present in the data. A graphical inspection of the smoothed risk function per day (rightmost panel of Figure 2) and of the percentages of events for unit at risk (not shown here) are more in favour of the raw estimate of day 75 as the end point based on a 20 days grid.

## 6 Discussion

We have developed a new procedure for estimating a change point in an L-shaped hazard function. We assume a declining hazard rate for the time interval preceding the change point $\tau$, and a constant hazard rate thereafter. This shape of the hazard rate is typical for applications in medicine when modelling survival of severely ill patients, and in material science for testing the reliability of complex electronic systems. Since we do not make any parametric assumption on the hazard function for the time interval preceding the change point, our procedure is nonparametric. While most approaches in the literature assume a jump in the hazard rate, our procedure is valid for data with and without a jump.

Our new estimator is based on p-values, which result from testing the hypothesis of a specific value for the hazard rates in suitable time intervals. The change point is estimated by fitting a stump regression model, with a jump at the change point, to these p-values. Our method was inspired by Mallik et al (2011), who adopt a similar strategy in a regression context and proofed it to be consistent. Furthermore, it represents the extension of previous attempts which, though plausible, did not have a theoretical foundation. These solutions were also based on p-values, but used the first significant p-value as estimating criterion (see our preliminary results in Schneider et al 2010).

Alternative approaches for finding a change point in a setting with two or more constant hazard rates (see e.g. Matthews and Farewell 1982; Loader 1991; Noura and Read 1990) are not appropriate in our situation where the hazard rates may rapidly decrease before the change point is reached. Maximum likelihood and pseudo maximum likelihood approaches using a Weibull model have been proposed in reliability theory to model the hazard rate for the time interval preceding the change point; see Kleyner and Sandborn (2005) and Altun and Comert (2016). These authors show that the parametric approach is useful for estimating the change point and for reliability prediction based on the respective model. However, maximum likelihood estimation may not be robust to model misspecification when estimating the change point; see Gürler and Yenigün (2011). In our applications, the different maximum likelihood approaches did not lead to acceptable results. In our simulation study, we found that the maximum likelihood estimator is highly non robust, i.e. small deviations from the distributional assumptions or rounding led to high bias.

An estimation strategy, which does not rely on parametric assumptions and looked appropriate for our setting, has been proposed by Yang et al (2012). These authors develop a rather complex Bayesian procedure, for which however no software is provided. Hence, we weren't able to compare their strategy with our new approach.

We checked the performance of our new estimator by an extensive simulation study. In scenarios, which are similar to our data applications, our new estimator showed a good performance, especially for those set-ups with a drop in the hazard rate or with a rather steep decline of the hazard function in the

time interval preceding the change point. In particular, the bootstrap based confidence intervals had an acceptable coverage rate. Furthermore, our estimator reliably identified the change point in data sets with varying contents of information.

Our new approach has also some limitations, which derives from the necessity to define two tuning parameters (see Section 3.1). First, we need to fix an upper bound for the change point (Step 1). In our applications it was always possible to identify reasonable values for $\tau_{max}$ by graphical inspection of the smoothed risk function, and our estimation procedure was robust concerning the choice of this upper bound. Second, the grid length to be used in our procedure (Step 4) is a tuning parameter, which, according to our simulation studies can have a relevant effect on the performance of the estimator. In Section 3.3, and for the real data applications, we carefully discussed this aspect. Furthermore, we developed and illustrated some graphical tools to support the decision on which value to choose (Section 5). The new methodology presented in this paper was implemented in the R-package called `CPsurv` (*Nonparametric Changepoint Estimation for Survival Data*), which is freely available on CRAN. A data driven choice of the tuning parameter is subject of further research, as well as the formal proof of consistency for our estimator.

Our new estimator proved to be highly effective in medical applications. In Section 5, we discussed three different applications in detail, which showed that our procedure is useful for handling difficult data problems. From a practical perspective, our results can be expected to have a relevant impact on healthcare performance measurements and on the definition of hospital performance metrics. In hospital systems benchmarking performance is an essential part of quality control, and mortality after a procedure or injury is one of the most important performance parameters. Usually, mortality data are based upon those who survive only to hospital discharge (Callcut et al 2016). There is, however, growing evidence that a significant portion of patients, who had a major procedure or injury, will die after discharge or transfer to other institutions such as rehabilitation or weaning units (Callcut et al 2016; Schneider et al 2010). These observations have led the *British Medical Journal* to declare that analyses based only on in-hospital deaths can be potentially biased by differences in hospitals' discharge practices (Pouw et al 2013). Therefore, Schneider et al (2010) proposed to expand the follow-up period used for benchmarking until a point, which is no longer defined by the location of the patient, but by his survival time. A reasonable end point of the follow-up period would be the day beyond which there is a relative stabilization in survival. For critically ill patients, it has been suggested to use day 60 or day 90 after ICU admission as the end point of the follow-up. These time spans, however, are subjective and have not been based on the empirical properties of the survival function. Therefore, an objective change point estimation procedure based on patient data—such as our—is highly desirable to standardize performance measurements.

## Acknowledgements

We would like to thank the Associate Editor and the two anonymous Referees for their careful reading of the paper and the most useful comments which greatly helped us improving it.

## References

Altun M, Comert SV (2016) A change-point based reliability prediction model using field return data. Reliability Engineering and System Safety 156:175–184

Antoniadis A, Gijbels I, Macgibbon B (2000) Non-parametric estimation for the location of a change-point in an otherwise smooth hazard function under random censoring. Scandinavian Journal of Statistics 27:501–519

Callcut RA, Wakam G, Conroy AS, Kornblith L, Howard BM, Campion EM, Nelson MF, Mell MW, Cohen MJ (2016) Discovering the truth about life after discharge: Long-term trauma-related mortality. Journal of Trauma and Acute Care Surgery 80:210–215

Chang IS, Chen CH, Hsiung CA (1994) Estimation in change-point hazard rate models with random censorship. In: Carlstein, E and Müller, H G and Siegmund, D (ed) Change Point Problems, Institute of Mathematical Statistics Lecture Notes, vol 23, pp 78–92

Davison AC, Hinkley DV (1997) Bootstrap Methods and Their Applications. Cambridge University Press, New York

Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap. Chapman & Hall, New York

Eriksson M, Brattström O, Larsson E, Oldner A (2016) Causes of excessive late death after trauma compared with a matched control cohort. British Journal of Surgery 103:1282–1289

Gijbels I, Gürler U (2003) Estimation of a change point in a hazard function based on censored data. Life Time Data Analysis 9:395–411

Gürler U, Yenigün CD (2011) Full and conditional likelihood approaches for hazard change point estimation with truncated and censored data. Computational Statistics and Data Analysis 55:2856–2870

Kleyner A, Sandborn P (2005) A warranty forecasting model based on piecewise statistical distributions and stochastic simulation. Reliability Engineering & System Safety 88:207–214

Li Y, Qian L, Zhang W (2013) Estimation in a change-point hazard regression model with long-term survivors. Statistics & Probability Letters 83:1683–1691

Loader CR (1991) Inference for a hazard rate change point. Biometrika 78:749–757

Mallik A, Sen B, Banerjee M, Michailidis G (2011) Threshold estimation based on a p-value framework in dose-response and regression settings. Biometrika 98:887–900

Matthews DE, Farewell VT (1982) On testing for a constant hazard rate against a change point alternative. Biometrics 38:463–468

Müller HG, Wang JL (1990) Nonparametric analysis of changes in hazard rates for censored survival data: An alternative to change point models. Biometrika 77:305–314

Noura AA, Read KLQ (1990) Proportional hazards change point models in survival analysis. Journal of the Royal Statistical Society Series C (Applied Statistics) 39:241–253

Pouw ME, Peelen LM, Moons KG, Kalkman CJ, Lingsma HF (2013) Including post-discharge mortality in calculation of hospital standardised mortality ratios: retrospective analysis of hospital episode statistics. British Medical Journal 347:f5913

Schiergens TS, Dörsch M, Mittermeier L, Brand K, Küchenhoff H, Lee SML, Feng H, Jauch KW, Werner J, Thasler WE (2015) Thirty-day mortality leads to underestimation of postoperative death after liver resection: A novel method to define the acute postoperative period. Surgery 158:1530–1537

Schneider CP, Fertmann J, Geiger S, Wolf H, Biermaier H, Hofner B, Küchenhoff H, Jauch KW, Hartl WH (2010) Long-term survival after surgical critical illness: the impact of prolonged preceding organ support therapy. Annals of Surgery 251:1145–1153

Wang J, Zheng M, Yu W (2014) Wavelet analysis of change points in nonparametric hazard rate models under random censorship. Communications in Statistics: Theory and Methods 43:1956–1978

Yang CH, Yuan T, Kuo W, Kuo Y (2012) Non-parametric Bayesian modeling of hazard rate with a change point for nanoelectronic devices. IEEE Transactions 44:496–506